Department of the Environment

1991 DEPRIVATION INDEX: A REVIEW OF APPROACHES AND A MATRIX OF RESULTS

1991 Deprivation Index: a review of approaches and a matrix of results

Part 1 Towards an index of deprivation: a review of alternative approaches

Mike Coombes, Simon Raybould, Cecilia Wong and Stan Openshaw
Centre for Urban and Regional Development Studies, University of Newcastle

Part 2 A matrix of deprivation in English authorities, 1991

Brian Robson, Michael Bradford and Rachel Tye
Centre for Urban Policy Studies, University of Manchester

This publication contains reports from the two contractors commissioned by the Department of the Environment to advise and work on an index which measures levels of relative deprivation across England.

The first report, 'Towards an Index of Deprivation: a Review of Alternative Approaches' is by a team from the University of Newcastle and was commissioned prior to results from the 1991 Population Census becoming available. (Cecilia Wong has since moved to the Department of Planning and Landscape at Manchester University, and Professor Stan Openshaw to the Department of Geography at Leeds University.) It provides a conceptual basis for the measurement of deprivation and puts forward some proposals for indicators. The report considers the value of Geographic Information Systems in compiling data, and reviews alternative statistical approaches for creating an overall index.

Subsequently a team from the University of Manchester were contracted to take this work forward and construct the index. A decision was taken at this stage to broaden the choice of possible deprivation indicators by testing, not just those recommended by the earlier research, but others where data availability meant they could be used only at the local authority scale. Indicators went through a range of tests and two rounds of consultation took place with the local authority and voluntary associations about the proposals.

In May 1994 key results from the University of Manchester index were released by the Department of the Environment under the title 'Index of Local Conditions'. The contractor's full report, 'A Matrix of Deprivation in English Authorities 1991', is now complete and forms the second part of this volume.

Page

Part 1

Towards an index of deprivation: a review of alternative approaches

CONTENTS

Acknowledgements

Executive summary

Deprivation continues to pose a major challenge for urban policy in Britain. In the present policy context, a key concern is the targetting of assistance on the greatest concentrations of deprivation. This report seeks to identify the best approach to using the 1991 Census data for targetting the areas of most acute deprivation.

The background for the study was provided by the analysis of 1981 Census data which the Department used in its designation of Urban Priority Areas. This analysis has been widely quoted and also used by other organisations in their response to deprivation. Through the 1980s there has been much discussion of the nature of deprivation and also the best ways of identifying it 'on the ground' – but a consensus has not emerged on either issue. This report briefly reviews both debates and recommends a possible best practice for targetting analysis.

The nature of deprivation needs to be seen within a broad perspective, so as to include not only poverty and its consequences but also social and environmental problems. These latter, neighbourhood-level, problems fit very readily into an *area* targetting analysis. This report identifies ten *issues* in its anatomy of deprivation: social environment, physical environment, housing, education, employment, work conditions, income and needs, communications, recreation and health.

One of the over-riding principles which emerges is that the targetting analysis should focus on actual outcomes rather than on vulnerable groups. Many people belong to one or more vulnerable group (eg. the elderly, or ethnic minorities) but are not deprived. The need is for objective measures of deprivation 'outcomes' – such as the level of unemployment – whether or not the people concerned perceive themselves as being deprived.

Moving on to identify these objective measures, a set of criteria are required to select the appropriate variables. The report recognises that these criteria can conflict and so they need to be listed in order of importance:

1. robustness – each measure to be statistically sound and not unnecessarily complex;

2. relevance – each to provide a recognisable portrayal of one or more deprivation issues;

3. flexibility – every measure to be distinct so that its value to the analysis can be assessed separately;

4. hierarchy – each measure to be available at both the neighbourhood and the local authority levels;

5. dynamic – measures which were available in 1981, and can be updated by 1996, are preferable.

The eventual selection of *indicators* – statistical measures which meet the above criteria – had to greatly play down the fifth criterion because it would rule out most of the data which meets the other criteria. For this study, then, the question of measuring change in levels of deprivation was set aside as an issue to be considered last.

Of the remaining criteria, the need for a robust measure to be available at the neighbourhood level was most often the severest constraint. Two of the ten issues – work conditions and leisure – could not be represented by any indicator which met the criteria. Twenty-one indicators were selected as promising measures of various aspects of the other eight issues.

Some of the proposed indicators use previously unavailable information, while others rely upon new forms of analysis such as Geographic Information System (GIS) techniques. The study recommends a preliminary stage of *validation* analysis with the 1991 Census data. New indicators need to be assessed for their robustness and relevance, while the whole dataset will need some standardisation and testing for any unintended 'double counting' which can result from indicators duplicating each other.

The eventual set of indicators can be combined into a single index of deprivation in a number of different ways. The report considers and rejects two broad approaches, one of which treats the indicators as equally influential, and the other which 'weights' each one on a pre-determined basis. The general approach which is recommended relies upon statistical analysis of the indicators in combination, to be undertaken as the last stage of the validation analysis.

Five types of statistical analysis are reviewed, with the most simple being that used by the Department's 1981 analysis. Simplicity is an important advantage, but other limitations of this 'default' method (Z-scores) lead to the recommendation of factor analysis as the most promising alternative. However, in the validation analysis one of the other methods may prove to be better than factor analysis at capturing the variation within the 1991 data.

Returning finally to the question of measuring change in deprivation, there is no wholly satisfactory option for subsequent updating of the recommended 1991 index. The validation analysis will again provide important information on the best way forward. As for 1981-91 change, the 1981 index should be updated with 1991 data.

1 Introduction

1.1 The problem of deprivation is now an all-too-familiar feature of contemporary Britain. Much has been written on the nature of deprivation, who suffers it, and why. Many public policies have addressed themselves to the problem, often taking the inner city as a particular focus for their actions. The tone was set by an earlier White Paper which asserted "inner areas of cities have a higher concentration of poor people" (House of Commons 1977). This report is the first step towards an analysis which would assess whether deprivation is still concentrated in the inner cities.

1.2 Despite the continuity of concern (cf. Lawless, 1979), there is no doubt that deprivation continues to pose difficult challenges. In seeking to respond to deprivation, policy makers need to understand the nature and roots of the hardships suffered by those who are deprived. In the current policy context, they also need a means of targetting areas in which deprivation is concentrated. The term targetting is used to stress the geographical basis for the analysis, rather than to imply any direct resource implications from this study. This report provides a review of targetting analyses, in advance of the 1991 census data becoming available for a major update of the 1981-based deprivation analysis (Dept. of Environment, 1983).

1.3 There are five critical elements to a targetting analysis (Coombes et al, 1994):

(i)   clarify the concept to be expressed

(ii)   recognise the principles for the targetting analysis

(iii) specify relevant and robust indicators

(iv)   devise an index which synthesises the indicators

(v)   implement the analysis, while evaluating the initial results to allow for late revision and improvement.

1.4 Each of these elements is the subject for one of the following Sections (2 to 6) of this report. The first and most fundamental task is to specify the concept of deprivation which will provide the underpinning of all the statistical work. At the time of the 1981-based analysis, the principal concern was with multiple deprivation. In this study, there is more discussion of the distinction between subjective and objective approaches – and also between measures of the prevalence of some problems, as against the concentration of groups who are particularly at risk. These discussions lead up to the identification of ten **Issues** which, in combination, provide a broadly based understanding of deprivation.

1.5 The third section of the report begins the process of translating the conceptual discussion into a targetting analysis. The over-riding constraint in that process is the need to analyse areas rather than people or households. This is such an important point that it is inappropriate to simply proceed as if the analysis was of microdata, but 'at one stage removed' in practice. For example, in a spatial analysis it is rarely possible to distinguish between a measure of prevalence and one of risk. This section of the report re-casts the objectives for the analysis from those that were implied by the conceptual discussion, into a series of operationalisable principles.

1.6 The fourth section of the report moves on to the more familiar question of which statistical indicators are most appropriate. The discussions in the preceding sections of this report lead to a distinct approach to indicator selection (eg. explicitly considering 'neighbourhood effects'). Ranging across the ten broad **Issues** related to deprivation, several new possibilities are considered alongside many 'traditional' measures of deprivation.

1.7 The wide-ranging approach to indicator selection sharpens the question of how to combine several measures into a single targetting analysis. The fifth section turns to this question and thereby reviews alternatives to the particular approach used by the 1981-based analysis. This leads on to considering the methodological strengths of numerous highly contrasting approaches. Some of these would readily allow several 'variant' indices to be generated. The sixth section of the report stresses that the next stage of this analysis must be an assessment of the proposed indicators when 1991 data comes available. As a preliminary step, an experiment has been undertaken with 1981 data to assess the extent to which the different methods tend to produce different results.

1.8 The penultimate section of the report turns to the need for analyses of trends, to set alongside the targetting analysis that will be based primarily on 1991 census data. First, there is the question of reconciling the new measures with the 1981-based analysis – to identify how far their differing results are due to changes in the form of analysis, rather than to changes in relative levels of deprivation across England and Wales. Second, the desirability of some (partial) updating to the 1991 analysis, without waiting for another census of population, prompts a discussion of the extent to which updating will be feasible.

1.9 The report ends with brief conclusions on the strengths and weaknesses of its recommendations. Some remaining uncertainties are highlighted, particularly in relation to timing and census availability. Even so, the key points can still be rehearsed within the framework of the study, which focusses on the six problems – the concept, approach, indicators, methods, need for verification and of updating – which are inherent for a targetting study.

2 Deprivation: towards an operational definition

2.1 Although there is no shortage of academic and policy debate around the issue of deprivation, there is no single definition of the concept which commands wide acceptance. A number of reasons contribute to this continuing uncertainty:

- deprivation is politically sensitive and as such raises questions that are difficult to resolve in purely 'scientific' value-free terms;

- deprivation is often used in compound terms (eg. 'social deprivation') whose definition *may* be more precise, but which then make it more difficult to find a definition for the general term of deprivation that is applicable across all of these more specific versions; and

- deprivation has a battery of (near) synonyms – poverty, need, disadvantage, insufficiency, underprivilege – that confuse the task of finding a precise definition that may apply to deprivation but not include all aspects of the related concepts.

2.2 The fundamental implication of the term deprivation is of an absence – of essential or desirable attributes, possessions and opportunities which are considered no more than the minimum by that society. Surprisingly few attempts have been made to provide a comprehensive list of these desirable features of which a person can be deprived. Before turning to this question ("deprived of what?"), there are some more basic questions to be addressed.

Deprived according to whom?

2.3 Perhaps the most fundamental issue here revolves around a doubt over the standpoint from which deprivation is identified. The stark distinction is between subjective and objective: whether a person's deprivation is to be assessed by themselves or in a more detached way. Clearly, there will be examples of people who consider themselves deprived because they crave some asset or attribute which, however, is not considered to be essential by most members of that society. Conversely, other people may be lacking commonly-accepted necessities, yet not feel 'in need' of them at all. Mack & Lansley (1985) developed an approach based on asking individuals what they think is generally "necessary" and then deriving the social norm from these replies. The next step is an objective measurement of individuals' circumstances against this set of standards. In this way, the measure of deprivation is essentially objective – it doesn't ask if people feel deprived as a result of their own circumstances – yet the measures can also be responsive to changes in society, rather than trying to define a minimalist set of "basic needs" which would not reflect the wider society's aspirations (Townsend 1987).

5

2.4 There do remain a number of questions here – is it possible for a form of deprivation to be afflicting (part of) society without the issue being acknowledged by many members of that society – and how many need to assert that something is "necessary" before it is said to be generally accepted? The constraints upon this study, requiring an analysis at the very local scale across the whole country, make it self-evident that a purely subjective approach is not feasible. The analysis here will be restricted by the availability of information from secondary sources, so conceptual issues can rapidly decay into questions as to what has already been measured for other purposes. In any case, the research needs to avoid any approach which is too sensitive to the social norms of the moment – or of a particular region or social group – because it is to be applied across the country and is expected to remain relevant for up to a decade. Taking a strictly objective methodology, however, still leaves opportunities to adopt an expansive approach to the definition of deprivation.

Who or what is deprived?

2.5 One of the most important underlying issues has in fact been set aside in the preceding section: that is, the subject of the analysis. Is deprivation suffered solely by individuals, or perhaps by households, or even by communities? Once again, the imprecision of the term deprivation is involved in this question. It would be possible to define deprivation exclusively in terms of the individual's experience, but this could rule out of consideration a range of social and environmental concerns which are prominent in many authorities' understanding of deprivation. On the other hand, early studies based entirely on Census data tended to identify "deprived areas" – and in so doing lost sight of the fact that these areas included many people with few if any problems, and also excluded many of the most deprived individuals because they were scattered across areas where most people were more fortunate (Holterman, 1975, Brown & Madge, 1982). This brings back into focus the practicalities of measurement because, even if the *concept* is defined exclusively at the individual or household level, almost all secondary source data is for areas.

2.6 One way into this question is via examining a particular problem such as unemployment. The available information will be in a form which allows an area's unemployment rate to be calculated. To say "the area suffers from a 40% unemployment rate" is clearly a short-hand form of description. It is not a claim that there is a constant level of deprivation across the area and its inhabitants – in contrast to, say, a measure of air pollution which may genuinely be almost constant across the area. In fact, no-one is suffering at the 40% level: 40% of individuals are suffering at the 100% level. Even within one household, the experiences of different individuals will vary between the two extremes (0% and 100%). Of course, the area measure should be seen to be one of *prevalence* of a problem experienced by individuals – but it may also represent a measure of *risk*. Some of the influences upon an individual's likelihood of suffering unemployment are indeed localised (most notably, the availability of jobs) so the local level of joblessness can also be seen as a measure of risk to which all local residents are exposed (Daniel, 1990). The same argument is likely to apply to a number of other problems, such as crime or road accidents.

2.7 The implication must be that the individual *is* the proper subject for a definition of deprivation, because only then is it possible to examine the different experiences

that arise from local risk factors impacting on personal and household circumstances. By the same token, the individual's experience will also be shaped by factors related to the household (eg. race and class) and the neighbourhood (eg. environmental and social conditions).

2.8 An increasing number of studies are responding to conclusions of this nature by adopting a multi-level modelling approach (eg. Goldstein, 1987). This form of analysis is undertaken at the individual level, but allows for the influence of different factors at different spatial scales. For this national study of deprivation, however, the absence of relevant individual level data is a mortal blow to any thought of adopting a multi-level approach. The more 'traditional' area-based targetting analysis could, however, be enhanced by linkage to extensive survey-based research which explores the relative influence of individual, household and local scale factors upon the experience of individuals. In this study, the analysis must be limited to areas so the measures have to be seen as the prevalence, or risk, of deprivation being experienced by the individuals in that area. Because of this limitation, it will not be possible to distinguish between measures of risk and of prevalence – the available data can be interpreted as either or even, perhaps both.

Actually or potentially deprived?

2.9 The last underlying question related to deprivation is whether it is possible, even if only in principle, to clearly separate deprivation *outcomes* from risks or conditions which are not necessarily forms of deprivation. The concept of deprivation would appear to emphasise outcomes: that is, the actual experience of individuals of households. Thus an individual could be said to suffer material deprivation if they are poor – an outcome, in many cases, of their own occupational status together with local labour market conditions. In contrast, these *conditions* which apply to the individual may or may not cause deprivation as an outcome in any individual case.

2.10 In practice, however, such a distinction is not sustainable. The notion of a 'cycle of deprivation' illustrates the problem: individuals who are poor are also more likely to live in unsatisfactory housing conditions and to suffer health problems, thereby endangering their employment status and thus reinforcing their poverty. In this way, each outcome is *also* a condition which makes the sufferer more vulnerable to other aspects of deprivation. For example, there is a burgeoning literature on the links between the social and physical environment generally, housing in particular, and other forms of deprivation (eg. Hatch & Sherrott, 1973, Kearns, 1990, and Carley 1990). The tendency for individuals to thus experience more than one form of deprivation has been simplified in the term multiple deprivation.

2.11 The inclusion of risk factors brings with it an interest in a wide range of social and environmental concerns. Deprivation is not limited to material well-being: if the term "poverty" were used then it would have to be extended to include, for example, poverty of social life and access to amenities. One related approach focusses on "resources" by suggesting that deprivation is the lack of the many and varied resources that can be needed to obtain the quality of life that is expected in our society. However, this approach is essentially a re-labelling of the attempt to distinguish conditions from outcomes and as such is not really helpful.

2.12 However, not all the conditions which are likely to result in deprivation are, in themselves, forms of deprivation. For example, vulnerable groups (such as the elderly or handicapped) should not necessarily be seen as being deprived simply because they have a higher risk of one or more form of deprivation. Given this study's emphasis on outcomes, data on a vulnerable group should only be used as a 'proxy of last resort' where there is no available data on outcomes of the particular form of deprivation to which that group is vulnerable. The difficulty of distinguishing prevalence from risk would thus be exploited by using measures of the latter to 'predict' the former: however, the intention here remains to focus purely on outcomes, which will also allow subsequent analysis to assess empirically which groups continue to be vulnerable to which forms of deprivation.

Deprived of what?

2.13 So far then, the definition of deprivation has been narrowed down to an objective assessment of individuals' circumstances, measured at the area level against social norms. On this basis, it is neither possible nor really justifiable to seek a distinction between the prevalence and the risk of deprivation in a neighbourhood – except to avoid measures which focus on vulnerable groups who are not, in themselves, deprived. In reaching these conclusions, a broadly-based notion of deprivation has been presumed. Whereas there has been an active debate over the broader questions discussed earlier, Townsend (1987) is one of remarkably few detailed attempts to systematically categorise the relevant aspects of deprivation: that is, on what issues can a person be said to be deprived?

2.14 Smith (1979) suggests that one important starting point for a generalisable approach is the United Nations' "Level of Living" (LofL) components (UNO, 1954). These components were designed for broadly based assessments of variations in experience across time and space. The dozen LofL components are: human freedom, social security, housing and household facilities, education and skills, employment status, conditions of work, consumption and saving, food and nutrition, transport, recreation and entertainment, health and population factors.

2.15 The last category in the LofL list is essentially similar to the vulnerable groups which are to be set aside from this study. Otherwise, the LofL components provide a valuable checklist which seem as relevant now as in 1947 when they were devised. The one major issue which has risen to prominence subsequently is probably the concern over the environment. Following on from this observation, then, **Table 2.1** takes the LofL list and translates it into a series of **Issues** for this study. For example, the first two LofL components can be more appropriately interpreted together as Social Conditions, given this study's focus on small areas within a single nation (as opposed to the UN's concern with differences between states).

How can the definition be operationalised?

2.16 An initial response to **Table 2.1** could be that the **Issues** there seem likely to command general agreement. Yet moves towards an operational definition of deprivation often arouse controversy. A particularly active debate is currently underway on how to measure deprivation, as one element underlying patterns of inequality in health (Townsend et al (eds.) 1988). The influence of deprivation is now generally accepted (Whitehead, 1987) and statistical evidence for the

association with ill health has been established (Townsend, Phillimore & Beattie, 1988). However there remains substantial scope for alternative deprivation measures to be tested for their statistical strength of association with ill health (eg. Campbell et al, 1991 and Morris & Carstairs, 1991). The key point here is that there can be a weakness in trying to operationalise the conception in terms of a measurement. In the health field, it may be that poverty is the critical issue, but this could not be operationalised because accurate local poverty data is virtually non-existent in Britain (Townsend & Gordon, 1991). On the other hand, ill health may be an outcome of a much more generalised syndrome of deprivation, whose key dimensions are not yet clear so the need for an operational definition simply leads to a competition between statistical analyses.

Table 2.1   Issues of deprivation: the UN Level of Living components

UN Level of Living component	Deprivation Issue DEPRIVATION OF....	ISSUE CODE:	Related concerns
Human freedom	SOCIAL ENVIRONMENT	A	alienation
Social security			
	PHYSICAL ENVIRONMENT	B	pollution
Housing & household facilities	HOUSING	C	homelessness
Education & skills	EDUCATION	D	lacking skills
Employment status	EMPLOYMENT	E	unemployment
Work conditions	WORK CONDITIONS	F	hazardous jobs
Consumption & saving	INCOME AND NEEDS	G	poverty
Food & nutrition			
Transport	COMMUNICATIONS	H	inaccessibility
Recreation & entertainment	RECREATION	I	lacking leisure
Health	HEALTH	J	mortality/illness
& population factors	< vulnerable groups >		ethnic minorities

2.17 The very broadness of the concept of deprivation – as illustrated in **Table 2.1** – makes a sharp definition very difficult to achieve. The very patchiness of the available data for measuring deprivation makes sharp definitions sterile – the only plausible candidate for a strong 'lead' indicator is income, and data on this is not available for small areas in Britain. Hence the resolution of this dilemma tends to be different for the particular purpose of each study, with one definition placing more emphasis on conceptual rigour and another on analytical precision.

2.18 It is perhaps useful to begin the conclusions to this section by emphasising those points which are most likely to attract a consensus. First, deprivation is most directly experienced by individuals, although it can be seen to operate at many levels. The need here for a targetting analysis implies an operational definition which uses data for areas, which in turn implies an objective rather a subjective notion of deprivation. The available small area data for aspects of deprivation is profoundly inadequate, with the most serious problem being the lack of data on household incomes. With this background, then, lengthy debate into the underlying processes of deprivation are unlikely to help in the choice between one imprecise statistical indicator and another.

2.19 The response here has therefore been to adopt a broadly-based framework for the targetting analysis. Ten major **Issues** have been specified, in order to exclude only those contributory factors to deprivation which make people vulnerable without being forms of deprivation themselves. The statistical corollary of this approach is to be open to the inclusion of numerous indicators, assessing empirically their possible contribution to an overall analysis which seeks strength in depth. The alternative which is rejected is that which asserts that just one or two issues need to be measured, and then adopts the best single indicator for each, regardless of data inadequacies. Given the known failings of the available small area data, the more robust approach is to seek a commonality of evidence across numerous indicators – in the belief that the very pervasiveness of key syndromes such as poverty will be identifiable in the recurring patterns of inequality which will emerge when the data on different aspects of deprivation are brought together.

3 Approach: objectives and constraints for this study

3.1 The preceding discussion has come to the conclusion that an anatomy of the concept of deprivation can only begin to sketch out a framework for the targetting analysis which is required here. Given the difficulty of precisely operationalising this framework with the available statistics, it is necessary to check for constraints arising from the policy context for the study. The first point is that the framework which has been set out does not severely clash with the basis for current policies. For example, the Training Agency (1990) recently identified many of the **Issues** of **Table 2.1** as part of its "challenge" – which had previously been found just in inner cities but is now also to be seen elsewhere. This study too rejects any restriction of its focus to the inner cities, not least because much recent research has shown that these areas no longer contain within them all the most deprived areas to be found when analysing issues such as youth unemployment (Garner et al, 1988). The area targetting analysis required here for policy implementation needs to meet the five objectives which were part of the brief for the study. The following discussion aims to resolve the inherent conflicts between the set of objectives.

Conflicting objectives

3.2 There are a number of objectives and requirements which the 1991 index of deprivation will be expected to meet. In particular, the aim of the research was the creation of an index of deprivation, based on 1991 information, which ideally would be:

(a) ROBUST — both statistically and in the sense of being readily understood and defended;

(b) RELEVANT — through reflecting the full range of factors involved in multiple deprivation;

(c) FLEXIBLE — in allowing for sensitivity testing of the results, and also supporting substantial changes to produce "variant" indices for different purposes;

(d) HIERARCHICAL — by virtue of being primarily defined at the level of the Enumeration District (ED) used in the 1991 Census, but also operating at higher levels; and

(e) DYNAMIC — with the potential for updating the results later in the 1990s, and to compare them with the 1981 analysis.

3.3 It is appropriate to start by considering the aim to relate the 1991 results back to the 1981 "Z-scores" analysis (Dept. of Environment 1983). Clearly, the *ideal* is not to just update the 1981 analysis with 1991 data (in order to give the most simple form of 1981-91 change measures). However it can be assumed that the 1991 index will differ from that of 1981 in form and content, not least to meet objectives **(b)**, **(d)** and **(e)** above. Fortunately the 1981-1991 change requirement can be met by re-running the 1981 Z-scores index on 1991 data in parallel with the new 1991 index. Consequently, whether or not it is possible (due to data availability) for the 1991 index to be backdated to 1981, *some* form of 1981-1991 change measure will be possible. As a result, the requirement for some form of 1980s change analysis should not be a critical constraint on the construction of the 1991 index.

3.4 In contrast, the *most* severe constraint could be posed by objective **(e)** – the aim to be able to update the analysis after 1991. This objective could be seen as directly conflicting with *all* the other objectives listed:

(i)   the data available in interCensal years is usually sampled, and as such would make the analysis less statistically robust (and also more liable to fluctuations year-on-year that would render the results less defensible);

(ii) the available nonCensus data will not cover all the issues in the Census (eg. single parent families) and this in turn will restrict the possibility for updating any variant indices that rely on Census-specific indicators; and

(iii) there is virtually no nonCensus data at the very local scale, and although various statistical techniques can be used to generate estimates at this scale these estimates are unlikely to often meet the requirements for robustness.

3.5 In short, to place too strong an emphasis on the need for mid-1990s updates would very severely restrict what could be done for 1991.

3.6 The approach here, therefore, is *not* to rule out any key feature of an 'ideal' 1991 index in order to ensure updatability – but where there are choices as to the components of that index, to favour any option which will ease the problems for updating subsequently. This report returns to the question of updatability in Section 7.

3.7 Of the remaining four objectives, **(a)** to **(d)**, robustness is clearly the most important. The statistical aspect of this objective will impose a 'conservative' approach towards the other three. For example, Geographic Information System (GIS) techniques will only be used to estimate data at the local scale if the indicator is known to vary predictably across local areas. In consequence, the index may not be able to cover all the relevant facets of deprivation. The more common sense aspects of robustness (that the results should be readily understood and defensible) also suggest a restrained approach to the flexibility objectives. This is because generating several similar indices – showing each area ranked in a slightly different way – may stimulate controversy.

3.8 Two of the remaining three objectives, **(b)** and **(c)**, are strongly related. The objective of relevance should in principle take precedence over that of flexibility: indeed, the ideal outcome could be that the final index is recognised to have been so successful in balancing together a wider range of deprivation issues that variant indices are not really needed. These two objectives together represent the *purpose* of the definitions, and this takes precedence over the slightly more technical concern of objective **(d)** regarding the areas of analysis. In the past, policy has not been operated at the very detailed level of the ED (which may have as few as 50 residents), although EDs were used as a stage in the analysis, reporting an area's "% of population in very deprived EDs" (Dept. of Environment, 1983). Doubts about the statistical robustness of analysing data at so fine a scale as EDs are inevitably heightened when the data is from the 10% sample in the Census. However, *any* set of areas will inevitably cause controversy in some locations when it is used in a policy context (eg. Dutton, 1992).

3.9 Thus the priority between the objectives should be that in which they were listed above, where **(a)** is highest. Even so, all the five objectives are potentially worth striving for – the next question is how they can best be interpreted as constraints for the subsequent analysis.

Practical constraints from the policy context

3.10 The index of deprivation to be generated following this study will be an important input to targetting policy activity. The 1981 index was used in many contexts, and attracted the suggestion of producing variant indices to reflect the different emphasis of the varied purposes for which it was used. Thus the flexibility objective **(c)** could be seen as raising the possibility of variant indices, each one targetting areas according to a different policy concern. This question is related to another one on the appropriateness of the ten **Issues**, identified in Section 2, from the viewpoint of policy: are some **Issues** very important in one area of policy but unhelpful in another? Consultation with Government Departments involved in urban and regeneration policies led to four conclusions on these questions.

(i) No respondent pressed for a variant index in advance of seeing the proposals for the 'general' index.

(ii) All respondents agreed that all the deprivation **Issues** put forward in **Table 2.1** were, at least potentially, relevant concerns.

(iii) No respondent was adamant that there were relevant concerns which lay outside these **Issues** (although the Department of Health respondent felt that family circumstances were so strongly associated with ill health that they were not just a "proxy of last resort" – and indeed could be seen as a form of non-material deprivation in many cases).

(iv) Each respondent suggested a different prioritisation between these **Issues** in terms of their particular Departmental emphases – but also agreed that the choice of indicators should not be driven by programme objectives – so it remains appropriate to develop a single 'general' index that each Department can set alongside their own programme indicators (and later perhaps allow users to adjust it interactively to develop purpose-specific variants).

3.11 In short, these responses pose no strong constraints on the remainder of the study. They do, however, suggest that there is an understandable concern to assess the output of the analysis 'on the ground' – and to keep open options for variant indices until the initial results have been assessed. Section 6 takes up the need for some initial analyses of the 1991 data before the targetting results are finalised.

Principles of indicator selection

3.12 The fundamental criteria for selecting indicators have been set by the prioritisation between the research's objectives, with the primary concern (constraint **(a)** above) that the eventual index should be statistically *robust*. The fact that the analysis will be conducted at a fine scale of spatial resolution underlines the requirement for high quality data inputs. Of course, many data sources are not made available for the small areas of interest here, precisely because those datasets depend on small samples which would be very variable in their accuracy within small areas. The eventual policy use of the index requires that the input data meets quite a high standard of precision in *every* area. Without doubt, the sheer level of its data collection ensures that the Population Census remains the most reliable source on these criteria. Even the Census, however, is vulnerable to the extreme values that are possible from the small samples found in very small areas. This risk is all the greater for variables depending on a 10% sample – including many of the valuable socio-economic variables – although the Census Office limits this risk by not providing the more vulnerable data for the smallest areas.

3.13 The principle to be adopted here is that additional sources should only be adopted where they add so much, in terms of information unobtainable from the Census, that this benefit outweighs the danger of weakening the robustness of an all-Census index. In general, the adding of indicators to an index may make it more robust because it will be less sensitive to any problems with one of the indicators. Yet there is also a presumption here in favour of simplicity, which will favour an index with fewer indicators. The 1981 index prompted queries as to 'double counting' in its use of strongly related indicators (eg. the *outcome* of unemployment as well as the *at risk* group of ethnic minorities). It will be necessary for the 1991 index to be supported by a 'validation' analysis of its statistical robustness, as discussed later in this report. The recommended indicators will need to be demonstrably valid as dimensions of deprivation: this will require analyses which cannot be undertaken in advance (eg. with 1981 data) because some 1991 Census variables were not present in 1981 (OPCS & GRO(S), 1992a).

3.14 A final point on robustness relates to the statistical form of the indicator. In many cases, there are options for the compilation of an indicator from the raw data. A clear case is provided by overcrowding, which in the past has been measured in the Census by the numbers of households of certain sizes in dwellings with different numbers of rooms. One option is to total the rooms and persons and produce a general occupation density measure. Another is to set a 'standard' of persons per room, and then measure the proportion of households on either side of this value. There are other approaches too: the selection needs to be by reference to the plausibility of the measure as a discriminator between areas of higher and lower levels of deprivation.

3.15 Turning to constraint **(b)**, the *relevance* of the indicator, the concern with plausibility shifts from statistical qualms to issues of interpretation. There is a common factor, however, in the impact of the spatial or ecological approach to identifying deprivation. If the analysis were at the household scale then there would be far fewer questions of either kind. For example, there is little doubt that however small an area, it will include households living at different degrees of under- or over-crowding. An ideal indicator would recognise all those whose living conditions are problematic, while also highlighting those in the most extreme circumstances. Given the limitation of the pre-defined Census tables, this has to be attempted with rather indirect measures. Any indicator is likely to run the risk either of focussing too narrowly on those in the worst circumstances, or of taking so broad a view that the problems of the genuinely deprived are submerged beneath information on the majority. The broader measures will only be valuable if the 'average' condition in any area provides a plausible surrogate for the size of the deprived minority there. In making such a judgement, the key evidence will be knowledge about the distribution of different groups of the population. For example, Berthoud & Kempson (1992) found that high levels of credit and debt were not characteristic of deprived areas, but that most of the *extreme* problems with credit *were* found in deprived areas. For this sort of problem, then, the 'average' experience of an area's population may not be a good guide to the probability of the area including more extreme experiences.

3.16 Thus the concern has shifted from statistical plausibility to interpretability. For example, the only direct evidence on educational attainment in the Census is on diplomas and degrees. Very few deprived people may have these higher qualifications, but there are many others without them who are not deprived. For our purposes, the question becomes whether the presence of highly qualified people living in an area is a good surrogate for the absence of altogether unqualified people. In statistical terms, this is clearly only true at the extremes (an area cannot have *very* high proportions of two sub-groups). However, general awareness of the social segregation of residential areas could suggest that the absence of highly qualified people as a reasonable proxy of the presence of unqualified people (as opposed to people with some basic level qualifications). Of course, measuring the absence of the 'privileged' (as a proxy for the presence of the deprived) would only be acceptable when there was no more directly relevant information available for an aspect of deprivation which was felt to be too important to be left unrepresented in the index. This judgement could become one aspect of *flexibility* constraint **(c)** which favours methods which can support different combinations of indicators. In practice, this probably also reinforces the preference for simpler variables, so that they may be more readily combined or separated for different purposes.

3.17 In the preceding discussions, there has been no mention of the *hierarchy* of spatial scales which represent constraint **(d)** for the analysis. The smallest possible unit of analysis is the Enumeration District (ED), with the Ward as the other possible basic unit in the Census. The ED provides a clear advantage in terms of spatial precision, and widens the hierarchical flexibility of analyses at different spatial scales. However it creates most difficulties in all other respects, with even the Census offering less information at this scale (and a higher likely level of error). It now appears to be unlikely that there will soon be, in all areas, a new 'default' set

of 1991 Census areas (between the Ward and the ED) which provides the finest scale at which all possible variables will be available (OPCS and GRO(S), 1992b). There has been a move away from defining special areas for Census analysis since the experiment with 1971 grid square data failed to gain wide support (cf. Coulter, 1978).

3.18 The fact that the ED is to be the unit of analysis here implies that each ED will often have to be set in its own wider context. In order to minimise the risk of placing too much reliance upon data from small samples, it will be sensible to base some indicators on a 'spatial weighted average' of the set of EDs within a given search area around that ED (see Martin, 1991). This should make the indicator more robust – and indeed is often plausibly related to the circumstances of the individual (eg. representing the wider area's housing opportunities). These factors can outweigh the disadvantage of reduced simplicity, because they allow recognition of genuine 'neighbourhood effects' such as poor access to certain facilities (eg. Coombes & Raybould, 1990). A secondary benefit of their wider spatial reference is that they will make the areas which are eventually defined as being deprived more likely to be continuous across space (ie. less fragmented) and over time (ie. less volatile).

3.19 This leads to the final, and lowest priority, constraint (e) – the *dynamic* analysis of change over time. In practice, the impact of the 'higher order' decision to favour Census data is to emphasise the value of measures which can be considered as essentially 'snapshot' information. Since this is the lowest priority objective, the implication must be that updatable information will only be preferable in cases where there is no Census data at all available, and there is no significant factor in favour of an alternative non-updatable dataset.

3.20 A related question concerns the possible inclusion of trend data within the index itself. One example of this approach is the "Booming Towns" series of reports (Champion & Green, 1990) which produces both a "static" and "dynamic" analysis, then a "combined" index that mixes the two sets of indicators. Introducing dynamic measures can reduce the options for the indicator combination method, because of their inclusion of negative and other less easily standardised values (eg. percent changes from an original value of zero). The assumption underlying this study is that change should be considered through a comparison of the 1991 index against its 1981-based predecessor. This comparison will be clearer if the 1991 index follows the 1981 practice of only using static data.

3.21 To reject change measures, however, could still leave open the option of including absolute values along with the more familiar ratio indicators. For example, the total number of unemployed might be thought to provide a distinct assessment of the extent of that problem, over and beyond the evidence obtained from an unemployment percentage rate. The problem is that such absolute values would mainly reflect the size of the unit concerned, and the definitions of these units are essentially arbitrary. The more fundamental point is that smaller areas often include a more extreme range of ratio values. However, the appropriate response is not to adopt absolute values instead, it is to keep this possible problem in mind when selecting the most appropriate form for any indicator.

3.22 The questions reviewed in this section have led to a series of guidelines for the selection of indicators in the next Section, where indicators will be sought which are:

- statistically robust;
- preferably based on 100% Census data;
- plausible and understandable;
- directly identifying deprivation;
- measuring outcomes and not vulnerability;
- at the ED level;
- preferably not estimated;
- static, but ideally updatable;
- ratio measures; and
- perhaps focussing on key aspects of the social or physical environment.

4 Measures: identifying indicators of local deprivation

4.1 It is now necessary to apply the principles of indicator selection. Each of the deprivation **Issues** are taken in turn, with a wide range of possible indicators considered against the principles. The main candidates were derived from an extensive review of the literature, which was drawn from a range of contexts. **Table 4.1** lists many of the candidate indicators from the literature, with an initial emphasis upon those in the 1981 index (Dept. of Environment, 1983). Contrast is provided by the Scottish equivalent analysis (Duguid & Grant, 1983), plus the well-known index of Jarman (1984) and one of the principal alternative approaches to it as a measure of the influence of social conditions on health outcomes (Townsend, Phillimore & Beattie, 1988).

A. Social Conditions

4.2 The concerns grouped under this **Issue** include those that were labelled "social malaise" in earlier studies (eg. Stewart et al, 1974). It is possible that a high level of residential turnover is often associated with social malaise – though this is clearly not always true, and indeed a complete lack of in-migration (the available Census measure) may also indicate a problem area. Although in-migrants are cross-tabulated against many other variables in the Census, no robust measure relevant to deprivation presents itself. One indicator of a local population's alienation which has been used in various contexts (eg. Morin, undated) is turnout rates at local elections – a measure which has the advantage of being updatable. In many cases this is likely to be a strong proxy for alienation, but it is probably too often affected by local circumstances to be consistently comparable between areas (eg. those wards which are political/marginal tend to be the wards where the electorate's interest in voting is highest).

4.3 Another tantalising prospect is an indicator of 'ghettoisation' – the concentration of one group to the exclusion of others. This can be measured quite simply for any set of groups – whether these be ethnic, social or lifestyle – with an indicator which suggests that *any* high level of concentration represents (potential) deprivation. However, each set of groups will include some more privileged group, and it is unlikely that an area with a strong concentration of this group is a 'deprived' area, simply by the exclusion of less privileged groups. Consequently the indicator would need to suggest that only areas with concentrations of the less privileged groups are deprived – in which case the indicator ceases to be a pure measure of 'ghettoisation' and becomes an indicator of vulnerable groups' presence.

4.4 Hope & Hough (1988) found that residents' general level of satisfaction with their neighbourhood was very closely linked to their perception of local crime

19

risk. Although there is a persistent concern with crime as the major feature of local social problems, many commentators (eg. Hough & Lewis, 1988, Evandrou, 1991) have identified the problems of relying upon officially collected crime statistics, especially when analysing data for areas within different Police Authority boundaries. The alternative data source proposed here is provided by the home contents insurance rates for each area, collated across several major insurers. This has been piloted in a recent study of Scotland (Coombes et al, 1994) and appears to closely mirror perceived levels of crime risk, which Herbert & Darwood (1992) found to be strongly related to evidence of burglary levels locally. Claims on insurance policies are used directly to set the rates for the following year, so the indicator is closely reflecting recent experience. The fact that there is a differing proportion of insured households in different areas is not directly relevant, as long as those who *are* insured accurately represent the crime risk experienced by the others in that area.

4.5 The insurance rate values are derived from the average experience of policy holders in each postcode district. A resident living near the edge of a postcode district is likely to be experiencing a less similar crime risk to someone living on the opposite side of that same postcode district, than they are to someone nearby but just across the district boundary. It is therefore appropriate to derive each ED's crime risk value by considering all the nearby districts' values and calculating a spatial average, weighted by the distance between the ED and the centre of each district [A1].

4.6 Rogerson et al (1989) found that fear of violent crime was the factor rated most highly in assessing quality of life, even though the experience of it remains quite rare. There is less need for a 'proxy' source of violent crime statistics because the official data (whether from the police, hospitals or the Criminal Injury Board) is less likely to be partial in its coverage than are the burglary statistics. However, The Economist (1990) suggests that there are only around 600 murders a year in England, so even adding other serious violence means that the rarity of the incidents makes them too sparse to reliably support an indicator at the neighbourhood level.

4.7 Another important feature of social conditions is community relations. The proposal here is to compare the unemployment rate of ethnic minority groups (in combination) against that of the white population [A2]. Green & Owen (1990) found that ethnic minorities were more at risk of long-term unemployment, as part of a set of inter-related problems which included evidence of discrimination against residents from particular neighbourhoods. There is the problem that other 'structural' factors (eg. age, or differing education levels) cannot always be standardised for with the Census data. However, if there *is* a dramatic difference in levels of education, say, then this too is likely to fuel alienation. In the case of the ethnic unemployment indicator, reliability will be enhanced by adopting the 'spatial average' technique (which is also justifiable by reference to labour market theories). The same techniques could also be used to check whether ethnic minorities are experiencing higher levels of overcrowding – or some other issue measured by the Census – but these could be questioned on the grounds of different cultural traditions having different expectations. In contrast, unemployment only measures these seeking work and failing to find it, so it appears to have the advantage of including an element of expectation – as well as actual outcome – in the data.

Table 4.1 Selected indicators from the literature

	DoE	TPB	Jar	SDD	Other Reference
A. Social Conditions					
crime rate					Lever et al 1990
local election turn-out					INSEE undated
residential turnover			•		Coombes & Raybould 1990
population change	•				Champion & Green 1990
B. Physical Environment					
road traffic accidents to children					Coombes et al 1994
proximity to 'noisome' features					Townsend 1987
environmental dereliction					Coombes et al 1992
pollution					Rogerson et al 1989
water quality					Fox & Goldblatt 1982
C. Housing					
no central heating	•			•	Newcastle CC undated
housing without amenities				•	Begg & Eversley 1986
households in mismatched housing					
overcrowding	•	•	•		Pilling 1990
level of homelessness					Berthoud 1983
D. Education					
low level of education (SEG/class)			•	•	Carstairs & Morris 1989
16-17 year olds not in education					Gray et al 1990
young out-migrants					Shucksmith 1990
labour force with degrees					Begg & Eversley 1986
E. Employment					
male unemployment					Carstairs & Morris 1989
access to job opportunities					Lever et al 1990
youth unemployment				•	
median duration of unemployment					Champion & Green 1990
total unemployment rate	•	•	•	•	Campbell et al 1991
economic activity rate					CEC 1991
F. Work Conditions					
deaths at the workplace					The Economist 1990
G. Income and Needs					
earnings levels per worker					Molho 1991
children in no/low earner household					Wrekin Council 1991
household below poverty level					Townsend & Gordon 1991
population obtain family credit					Pilling 1990
accessibility to shops					Coombes et al 1994
H. Communications					
households with no car		•			Carstairs & Morris 1989
telephone ownership					Newcastle CC undated
I. Recreation					
leisure time of non-dependents					Coombes et al 1994
availability of theatres etc.					Rogerson et al 1989
J. Health					
mortality rates/SMR	•	•			Lever et al 1990
SMR of pneumonia					Coombes et al 1994
permanent sick				•	Bell 1990

Notes:
DoE: Dept of Environment 1983
TPB: Townsend, P; Phillimore, P & Beattie, A 1988
JAR: Jarman, B 1984
SDD: Duguid, G & Grant, R 1983

B. Physical Environment

4.8 The main problem with the **Physical Environment** is that, unlike many of the human-based **Issues**, the various aspects of it are unlikely to be strongly correlated. Thus for other **Issues**, one very well-defined indicator may provide most of the relevant information on areas' relative deprivation on a range of inter-related aspects. For the **Physical Environment**, however, there is no clear link between such aspects as the dominant built form, the level of air pollution and the impact of noisome features such as airports. In practice, this problem is less taxing because there is little relevant data available at a fine scale. For example, it is not possible to implement at the ED level the indicator on the proportion of buildings pre-dating 1914 which was used in the original Inner Area classification analyses (of course, this might now be a less interesting indicator in any case, due to the recent growth of concern with some quite modern housing estates).

4.9 One important new source is digital map data. Each individual ED can be classified according to whether it is within 'nuisance' range of major features such as airports, motorways or railways. Given the improved quality data by mid-decade, other features such as sewage works or power stations could be added - although Ordnance Survey's costings are currently prohibitive. Existing low cost data (eg. from Bartholomew) is probably sufficiently accurate, although it cannot support detailed measures of proximity to beneficent features such as open space (which would also raise the question of whether deprivation was being defined to be urban, not rural).

4.10 Unfortunately, there does not appear to be an agreed list of identifiable noisome features which impact seriously on most local residents. Some people are deeply affected by living near dereliction (Coombes et al, 1992), but may be indifferent to other features which other residents abhor. Another challenge is the possible relevance to health expectations of drinking water quality (Fox & Goldblatt, 1982), but this would require detailed information on which source provides water to each neighbourhood. Information on other environmental problems, such as air quality, is also insufficiently targetted for present purposes. It is noteworthy here that the Department of Environment periodically collects information on derelict land, although the data has not previously been made available at a fine level of detail. It may be possible to process the detailed information into a suitable indicator using GIS (Wong et al, 1992), but as yet there is insufficient knowledge on how people respond to dereliction to be able to translate this into a precisely specified measure of the area 'affected' by each major derelict site.

4.11 One aspect of the built environment which has attracted some attention of late has been its implications for road accidents. For example, Lawson (1991) recognises that certain built forms and road layouts raise the risk of accidents to children in particular. There is nationally available annual data on road traffic accidents, coded by grid reference, which offers the possibility of subsetting to focus on accidents involving children. Walsh et al (1990) have shown that most accidents to children occur very close to home, so that the distribution of accidents can be used to generate a profile of the risk to local residents. One point raised by this measure [B1] is whether it would be better considered as an indicator of **Health** deprivation [Jn] – especially since accidents are the major cause of death among older children. This question may well be only one of labelling – but it would prove an important point if the eventual method of building an index was one

which required a specific number of indicators for each recognised **Issue**. The fact that there are other influences on the level of road accidents to children does not undermine the value of this indicator to the index of deprivation, nor does it negate the role of the physical environment in the accident data. However, there *is* clearly scope for indicators on other aspects of deprivation resulting from the **Physical Environment**, if appropriate information could be identified.

C. Housing

4.12 There is a clear link between this and the previous **Issue** – for example, both might be indirectly represented by measuring the share of local housing which is provided by flats (or, if the Census data for Scotland was matched south of the border, accommodation with access which is not at ground floor level). In fact, this **Issue** is the one for which the Census provides the largest range of possible indicators. This variety does not ensure that all the data which might have been hoped for is indeed available. An effort was made to obtain Census data on the 'roofless' population, but this met with very limited success. The official data on the homeless is collected as a by-product of administrative procedures, with the result that it is effectively limited to families with children – and even then is not comparable between areas because of authorities' varying implementation of the procedures. In consequence it cannot be considered a robust source for comparison across the country, even if it could be made available at the very local level needed here.

4.13 Owner-occupancy has been widely used as an indicator of the *absence* of deprivation – associated with the idea that those in social housing are increasingly the marginalised groups of society. Recent changes in the housing market have undermined this view, due to the significant number of households which have experienced financial stress after moving into owner-occupancy. In any case, one of the principles for indicator selection here is that a measure of the non-deprived may give a very misleading view of the pattern of deprivation.

4.14 A rather different problem is raised by a measure of housing vacancy, which might indicate that the local housing is unsuitable, leading to the prevalence of vandalism or other forms of physical decay. Once again the data sources make it difficult to focus closely enough on problematic cases, so as to exclude quite positive factors such as the purely frictional vacancies caused by new building, renovation or an active housing market. The other current debate on housing markets revolves around the concept of affordability. Unfortunately, this is now increasingly recognised to be highly problematic to define as an indicator for comparing areas consistently.

4.15 Turning to the built environment aspect, it is possible to advance on the long-established analysis of dwellings which lack basic amenities (exclusive use of indoor toilet and bathroom). The 1991 Census identifies households which lack all, or at least one, of these amenities or some form of central heating, while also distinguishing households in non-permanent accommodation. Although having one of these disadvantages may not constitute a form of deprivation in its own right, it did appear from the 1986 survey of Newcastle upon Tyne City Council (undated) that the proportion of an area's households (or people) who live with one or more of these disadvantages is an indicator **[C1]** which identifies areas of housing deprivation.

4.16 A more focused approach can consider the mismatch of housing to the area's residents. In the data for Scotland, this could include children on non-ground floor accommodation, as well as all-pensioner households without central heating. This approach could even have replaced the previous one, by summing together these groups likely to be in housing stress (other examples could be children in non-self-contained housing). Unfortunately, the cross-tabulations available for England & Wales are less complete, and also frustrate attempts to compile any consistent version of this approach (which in this case *would* be additive, so that 'double counting' has to be avoided). As a result, the proposed indicator [C2] simply targets a small number of crucial examples of housing stress and sums these as telling factors in their own right. There is a possibility of overlap with the previous indicator: this is a good example of the need for validation analysis with the 1991 data, with the presumption that if the two measures are very highly inter-correlated then the former is the more robust.

4.17 Another important concern is overcrowding (cf. Pilling, 1990). The preferred measure in housing allocation procedures identifies a 'bedroom standard' for each household, but this cannot be operationalised with the Census data available in England. The established Census indicator takes a person/room ratio and simply counts the proportion of households above this level. One difficulty is selecting the level at which overcrowding is deemed to begin: if this is high then an area with quite high occupancy may still have very few households qualifying to this level; if it is low then the really problematic cases may be swamped by information on far less crowded households. More of the original information would be retained by an ED-level ratio between all persons and rooms in the area, but this is even less focussed on the households with problems. However, if 'all rooms' are an adequate measure of housing provision, then a 'room standard' might be roughly developed in place of the 'bedroom standard' ideal. Unfortunately, it is not possible to follow the 'bedroom standard' principle of adjusting the required number of rooms according to the household's composition (eg. whether it includes a married couple). The proposal here [C3] is to identify those households with more than one person per room, and count the number of rooms they are short of this standard. This sum would then be expressed as a ratio over all households in the ED. The measure remains far from ideal, but appears to be better than a simple ratio in identifying the extent of an area's problem of overcrowding. However, this proposal should be compared empirically with the more familiar indicator, as part of the validation analysis with 1991 data.

4.18 The limitations of the pre-tabulated Census data might have led to recommending that special cross-tabulations were requested from OPCS. Unfortunately, user-specified analyses of the raw data will only be processed later, so the indicators could not be available within the standard Census publication schedule. After the verification analyses on the English 1991 data, and especially the 'room standard' measure – with comparisons against the Occupancy Norm data available in Scotland – some tabulations may need to be ordered to provide a context for the proposed indicator [C3]. It is possible that an improved indicator may become available for subsequent revision of the 1991 index, although it is important to note that OPCS will be cautious about generating previously unplanned variables at the ED level. One final possibility is that this form of special tabulation may prove to be valuable when the new English House Conditions Survey data

has been analysed. Linkage to the Census may suggest that analysing the Census data in some new way would provide a good proxy for the poor housing conditions which Carley (1990) stresses tend to be closely associated with poverty and deprivation.

D. Education

4.19 Section 3 of this report discussed the possibility of an indicator on the proportion of highly qualified people in each area. The distinctive residential preferences of some well qualified people may undermine any assumption that this 'privileged' group will be located in an inverse pattern to that of the least qualified group. A new proposal could be to derive a ratio of numbers in the early teenage group to those in their later 20s to identify the differential effect on youth migration (Shucksmith, 1990). This is clearly not simply an education indicator, but there is evidence to suggest that the most able and well qualified are the most likely to move. As a result, areas with fewer young people in their late 20s compared to those in their earlier teens (relative to the national ratio) are likely to have lost many of the better educated group. However, even deriving a spatial average on this indicator could throw up unpredictable contrasts between areas due to the tendency of many younger adults to live in inner cities. As a result, the proposal here [D1] is to use a socio-economic group (SEG) definition of the low-skilled to identify people who are unlikely to have many formal educational attainments. Although at the individual level some women's SEG classification is unreliable, at the neighbourhood scale there is no known systematic bias to the data.

4.20 In the near future it will be possible to compile exam results data for all schools. GIS processing would be needed to relate each school's data to the neighbourhood whose prospects it influences. These measures will not only reflect the performance of the schools themselves, but also "the handicaps imposed by the environment" which have been recognised since the Plowden Report (Central Advisory Committee for Education, 1967). The recent trend towards greater choice for parents means that a local school's records may no longer closely reflect local children's performance – nor their opportunities – because of the ease of choosing to use another school. A stronger indicator [D2] for this **Issue** is more likely to be the proportion of 16 and 17 year old residents who are in full-time education but not on work-related training schemes (Gray et al, 1990). Taking a local 'spatial average' should pick up the combined effect of a community's attitudes and the local level of provision.

E. Employment

4.21 A first option to consider is the use of the Employment Census data that is available every two or three years. The earlier decision to be wary of introducing dynamic measures within the set of indicators is reinforced here by this dataset's small sampling frame at local levels – which casts severe doubt on monitoring growth/decline in job opportunities at the very localised scale that is of interest here. Equivalent trend data on unemployment is less problematic, but raises the question of updatable information on the whole workforce in order to identify whether changes in the unemployment rates being monitored are essentially due to demographic rather than economic process.

4.22 A familiar census indicator on this issue is a female activity rate. A more aggregate picture can be obtained as an employment ratio, which contrasts those in work with those either unemployed or economically inactive (but not of a dependency age). The underlying assumption is that many of those who are inactive are in fact 'discouraged workers' who no longer actively seek work due to the lack of local job opportunities leading to "under employment" (CEC, 1991). The 1991 Census provides improved data on hours worked, so this indicator could be developed still further to be a 'work ratio' which identifies the number of hours work available per member of the area's potential workforce. Of course, it is unclear that higher levels of hours work, at the individual level, could simply be equated with well-being. However this analysis operates at the area level, so the indicator will be estimating whether there is much work available (and on terms and conditions which lead to the opportunities being taken up). For this measure to provide an appropriate local measure [E1] which accurately relates to the question of discouraged workers it may be necessary to exclude young people, who are more likely to be in education rather than working. As a new proposal, the indicator needs to be assessed within the validation analysis (where it could be contrasted with the results of a simpler employment ratio).

4.23 The (un)employment **Issue** is, like **Housing**, another which provides plenty of indicator options, and precedents of uncertain value. The next proposal here is to focus on unemployment among semi-skilled manual and less skilled non-manual SEGs. The main objective is to 'discount' the basic occupational structure effects, in order to highlight distinctly local problems (eg. 'addressism' or stigmatisation). These broad groups are potentially quite mobile between job types, so high levels of unemployment among them would be strongly indicative [E2]. Finally, the prospect of an occupation breakdown in the near future to the Department of Employment's frequent local dataseries on unemployment provides one requirement for updatability - although the need for updated denominators (the number of each group in the labour force) may prove to be an insurmountable hurdle. This problem does not apply to unemployment rates disaggregated by gender alone, because data is available to allow local estimates of men and women in the labourforce. However, there are doubts about the reliability of counts of female unemployment (Moore & Townroe, 1990) due to the variation in female activity rates and the influence of the benefit regulations on this data for women. Male unemployment rates provide an updatable indicator [E3] which, incidentally, has arguably come nearest to achieving acceptance as the single best indicator of area deprivation levels (cf. Carstairs & Morris, 1989).

4.24 One way of side-stepping the denominator problem with updatable unemployment data is to focus purely on the experience of the unemployed. The proportion of the current pool who have been unemployed for a long term is known to be related to deprivation but is not a particularly robust indicator (Green & Owen, 1990). The likelihood of an existing unemployed person leaving the register is a calculable measure that reflects the impact on the residualised members of the labour force of local levels of job opportunity (Daniel, 1990). The indicator [E4] will certainly have to be processed as a 'spatial average' to reduce potential data problems due to small numbers in some EDs, or such localised effects as a 'blitz' on false claimants.

4.25 A different concern is the relative accessibility of employment opportunities (Lever et al, 1990), over and above questions of transport provision which are discussed later. The workplace data from the Census does provide information that allows a median commuting distance to be calculated for each ED's workers in the 10% sample. It will be necessary to limit the analysis to lower-skilled groups, so as to standardise for occupation effects (as stressed earlier for unemployment). Unfortunately, this information will not be available until at least a year after the other Census data discussed in this note. The proposal **[E5]** is to use the 1981 data for the first version of the 1991 index, then use the 1991 commuting data as part of the first updating of the index.

F. Work conditions

4.26 The availability of information on this **Issue** is all the more straightforward for being limited. At a local level there is insufficient data to count serious injuries or deaths at the workplace, as in The Economist (1990), for example. The only plausible indicator depends upon national measures of accidents at the workplaces of each type of industry. Using the Employment Census data in each area, it is then possible to identify the proportion of jobs available, in and around any ED, which are in each type of industry. It is then a straightforward task to generate an overall assessment of the relative probability of hazardous work conditions for people seeking work in that area and those within commutable distance. Both sets of input data are updatable quite frequently, but would probably need to be based on several years' data to make the indicator as robust as possible. However, there remains the critical problem that the national level of hazard by industry is applied to all establishments/classified to that industry. It is probably not realistic to assume that an industry coding alone is a sufficient indication of the hazardousness of working conditions at any particular workplace – there are clearly dramatic variations between large and small workplaces, process plants and their firms' offices, and so forth.

G. Income and needs

4.27 There is not much doubt that the single most valuable indicator for a deprivation analysis would be a measure of per capita income. Hirschfield (1989) documents examples of the way such data is used in other countries' deprivation analyses – but it remains unavailable here at the local level, even though it is collected in some official surveys (Savage, 1992). A first alternative here could be to focus on evidence of disposable income. One approach could be to identify households who are likely to have meagre incomes once they have provided for their housing needs. The data from the 1991 Census will be an improvement on 1981 in identifying households which own their homes outright – who therefore are not at risk of re-possession and should also have less housing related costs. At the ED level the only potentially valuable cross-tabulation is with car ownership. These two major ownership variables could be combined so to provide information on the distribution of households with *neither* of these two major assets. However, to focus on those who no longer have a mortgage will have a strong age bias built into it because few people under 40 will own outright. On the other hand, to include all who are still buying would clearly bring in people who are now known to be in considerable financial difficulties due to housing market changes. There are also non-owner occupiers who are far from being deprived, so it does not seem possible to specify the indicator sufficiently well for it to provide a surrogate measure of households with very low disposable income.

4.28 The next major question is the extent of local residents' participation in paid employment. There is clearly a strong overlap here with the (un)employment **Issue**, although the change in emphasis can be encapsulated by changing denominators from the workforce to the whole population. Once again the preferred approach would be to cumulate evidence of deprivation across major age groups in the population (eg. children in non-earner households, solo unemployed persons and so forth). Unfortunately the 1991 data appears worse than the 1981 Census in providing no evidence of pensioners' previous work-related social group – a major loss, given that pensioners dependent upon state benefits are a principal poverty-risk group. The best alternative could be pensioners who do not owner occupy, but this is a questionable variable by which to compare areas with very different tenurial histories. To complete the doubts over this approach, it appears impossible to adequately target those dependant on purely part-time work in low-earning occupations (there is no part-time cross-tabulation with SEG or social class).

4.29 Another tack is suggested by the European approach (ISSAS, 1990) of identifying "adult equivalence" levels of 'need' for each household (the second and subsequent adults are 0.7 of the first adult, a child is 0.5). The cross-tabulations in the Census do allow a simplified version of this to be cumulated to the ED level, and then related to the number of earners as a 'dependency rate' **[G1]** that is more sophisticated than the usual simple form. Of course, no allowance is made here for the income rates of the earners – it is necessary to assume that phenomena such as 'DINKYs' are more frequently associated with high than low earners when comparing data on areas.

4.30 This is another **Issue** that is very strongly related to **Income**. In the light of the previous discussion's conclusion that it was not possible to consistently identify poverty-risk across all broad age groups, the proposal here is to simply focus on a 'family poverty' indicator (Wrekin Council, 1991). The 1991 data will provide a clear indicator of the proportion of children who live in no-earner households. It is recommended that the indicator should also cover households with just one part-time worker. Unfortunately, it is not possible with Census data to add those whose household head is working in a low earning occupation group. This indicator **[G2]** should be assessed in the validation analysis, perhaps with a contrast made against some local information derived by the Dept. of Social Security in the form of estimated average earnings by residents in each part of the country (Molho, 1991). That data relates to larger areas than individual EDs but can identify the prevailing earnings rate in the wider area within which the residents of an ED seek employment (Coombes et al, 1994).

4.31 Several local authorities, particularly in West Yorkshire, have investigated the spatial pattern of households receiving benefit under the Community Charge payment system. This information appears to conform to local knowledge of poverty in each area, and is fully postcoded so that ED-level analyses are possible. The new Council Tax systems are being implemented in the same locally-specific way as their predecessors, so that it will continue to be very difficult to collate data for the whole country - quite apart from the likely problem of variations in the data due to uneven benefit administration by different authorities. For both these reasons, a national source for data on benefit receipt is much preferable (eg. Pilling, 1990).

Fortunately, the postcoding of Income Support recipients is now virtually complete and outline agreement has been given to provide ED-level counts of children in these households [G3]. The indicator would depend upon eligible non-claimants being either very few or randomly distributed. Many past studies have preferred to include within a poverty line those households who have incomes slightly higher than the benefit cut-off (Townsend & Gordon, 1991). The spatial distribution of this marginal group seems unlikely to be very substantially different to those who *do* receive benefits, so their exclusion may not be too great a loss for a targeting analysis. As a new indicator, of course, assessment in the validation analysis will be important.

H. Communications

4.32 The concern under this **Issue** is with mobility and access to various types of opportunity. Coombes & Raybould (1990) experimented with 'access' indicators to identify areas which are relatively remote from, say, fresh produce shops, but these have not proved to be realisable across the whole country with the available information. Another concern could be with access to employment, for which the Census provides information identifying households with fewer cars than workers. From a broader perspective, the more familiar indicator is the access to *any* car: it seems questionable to label a two car, three earner households as deprived.

4.33 The other important aspect is the availability of public transport. No comprehensive information is available on local provision, so the proposal here is to generate information from the Census data on mode of commuting. Areas where few if any workers use public transport – especially those who do not have cars themselves – are likely to be areas where the local public transport provision is inadequate. This indicator could be maintained separately within the index, but the proposal here [H1] is to combine it with car availability data. On this basis, transport deprivation is only acute for carless households in areas of poor public transport.

4.34 A final option could be to pursue a substitute for the telephone ownership information which is unavailable from the Census due to a late decision to leave out that question. It seems possible that BT could provide the Department with the number of domestic subscribers per postcode unit – which could then be grouped up into EDs [H2]. Of course, this data depends on collaboration from BT, who would also need to gain data from the Hull telephone suppliers. This information should also be updatable annually, although taking several years together might provide a more robust analysis.

I. Recreation

4.35 This is the one **Issue** in which 'access to facilities' appears to have the most valuable contribution to make (eg. Rogerson, et al, 1989). The key data needed, of course, would be a fully postcoded (or grid referenced) list of facilities of a relevant type – ideally with some data on the quantity or quality of service offered at each location. The twin problems that arise are justifying the selection of any particular facility type(s) in principle, and finding adequate information source(s) in practice. Coombes et al (1994) could only identify adequate sources of information on golf courses, major sports centres and other tourist related facilities such as museums. The focus here on deprivation suggests that more basic facilities such as libraries

would be more appropriate, but no national comprehensive data source has been located.

4.36 A different concern is the level of leisure time enjoyed by different people. Survey work by the Henley Centre (1991) estimates the average "free time" of people within categories of age, family and employment status. It is possible to use Census data to identify the share of each ED's population within each of these categories, and then compute an overall 'score' for the ED. Coombes et al (1994) found that a simple version of this approach is dominated by the proportion of children and (especially) elderly in the ED's population, so the option here would be to look only at the non-dependent age groups. However, this indicator would again depend on an assumption that a national relationship (between economic activity and "free time") can be universally applied across the country. Taken together with the lack of sensitivity to the distinct experience of the unemployed, this weakness is enough to reject the indicator here.

J. Health

4.37 The 1991 Census provides an important new information resource in the local data on limiting long-term illness (LL-TI). One opportunity that the 1991 cross-tabulations offers is assessing the number of people in households with a LL-TI sufferer. However, this information would be difficult to interpret – a high value suggests many people are 'affected' – but a low value may reflect LL-TI people suffering alone. At the ED level, only very broad age banding of this data is available, but it will still be possible to estimate 'quality years of life impaired' from LL-TI tabulations. One concern is that the broad age bands will conceal systematic bias (areas with high LL-TI having a high average age of the group in each age band). An experimental assessment of this problem seems advisable - the alternative (though not fundamentally 'safer') is to simply derive LL-TI rates for, say, the oldest pre-retirement age group [J1]. As a new variable from the Census, a particular concern is the quality of the data, which will be evaluated by the Census Office in the next few months (Wormald, 1991).

4.38 Of course, the most extreme forms of affliction end in death, so that mortality data is an important source (and is annually updated). Unfortunately, it is not currently reported at the detailed level of EDs, and the necessary denominators are not yet routinely available below the District level. A very different possibility is created by the publication (OPCS, 1990), at the broad level of Health Authorities or Districts, of Standardised Mortality Rates for afflictions such as pneumonia which (among the under 75s) are thought to be very clearly related to environmental conditions. Coombes et al (1994) estimate the risk level in each ED by reference to its relative proximity to the centroids of Districts in that wider area: an ED located very near to one centroid will have an estimated value very close to that District's average. However, there is likely to be substantial intra-District variation which must be 'smoothed' out of the data which can be assembled in this way. It may then be preferable to widen the analysis to include deaths from all causes, on the assumption that OPCS will make standardised data available at the ward level [J2]. To avoid possible bias (eg. from the distribution of long stay institutions), the ED data will probably need to be calculated as a spatial average across all wards in and around that neighbourhood.

4.39 There has been much debate in recent years on the connection between deprivation and ill health (eg. Townsend et al (eds.), 1988, Whitehead, 1987, Townsend, Phillimore & Beattie, 1988 and Carstairs & Morris, 1989). Although individual analyses have provided different interpretations and emphases, there is a recurring emphasis upon factors such as low social class, unemployment, lack of car and so forth. The ideal response here might be to draw up a short list of such factors and identify the proportion of households with perhaps two or more risk factors. The cross-tabulations available in the Census frustrate this approach, but may offer one alternative. A number of such factors are available in 'pairs' (eg. households with no car and low social status). By identifying three such pairs, each of which is available in combination with the other, it is possible to cumulate the number of people in each 'pair' (ie. those suffering at least two of the risk factors). Clearly, this method will 'treble count' any people with all three risk factors – the number of which is not identifiable from the Census. So long as such a weighting of this trebly-at-risk group is considered reasonable, such an approach is justifiable. One possible advantage of this approach is that it will bring in to the index one to two risk factors – one parent families in particular – which have not otherwise been recognised as a distinct element in other indicators. The three factors that are proposed for combination here (see [J3] in the **Annex**) are one parent families, lacking a car, and not being owner occupiers.

4.40 All the 21 proposed indicators are listed in **Table 4.2** and are more precisely specified in the **Annex**. By way of an overview, there are two important observations to be made. The first is that the lack of *any* recommended indicators for **Working Conditions** and **Recreation** – although very disappointing – at least illustrates that the strict criteria here for indicator selection have been maintained. This may be contrasted with, for example, Sheldon & Moore (1968) who aimed to measure "the quality of leisure" – or Richardson & Grieve (1988) who sought indicators of susceptibility to each of the seven deadly sins! The second observation is that the complexities and uncertainties involved in selecting indicators help to explain just why it is that controversy continues to surround the whole process of finding adequate measures of deprivation (cf. Begg & Eversley, 1986). The emphasis placed here on robustness, together with the interest in experimenting with new techniques to allow the index to cover a broad range of **Issues**, leads to the need for some validation analyses once the 1991 data is available (see Section 6).

Table 4.2 Key characteristics of the proposed indicators

		R	S <see Key>	T	Updatability
A.	**Social Conditions**				
	A1: crime rate (proxied by insurance costs)	?	x	$	annual
	A2: differential ethnic minority unemployment	?	s	@	next Census
B.	**Physical Environment**				
	B1: road traffic accidents to children	?	x	@	annual
C.	**Housing**				
	C1: housing without full amenities	!	s	%	next Census
	C2: households in mismatched accommodation	!	s	%	next Census
	C3: room shortfall from 'standard'	?	s	%	next Census
D.	**Education**				
	D1: low level of education (proxied by SEG)	!	t	%	next Census
	D2: 16 & 17 year olds not in education	!	s	%	annual (LEA)?
E.	**Employment**				
	E1: hours worked per non-dependent	?	t	%	next Census
	E2: low skill SEG unemployment rate	?	t	@	DE Group?
	E3: male unemployment rate	!	s	%	monthly [$]
	E4: likelihood of ceasing unemployment	?	x	@	annual
	E5: access to job opportunities	?	w	@	SWS late 1994
F	**Work Conditions**				
G.	**Income & Needs**				
	G1: 'adult equivalence' dependency	?	s	#	next Census
	G2: children in non/low earner households	!	s	%	next Census
	G3: children in households on Income Support	?	s	@	DSS supply?
H.	**Communications**				
	H1: no car & poor public transport	?	t	#	next Census
	H2: telephone ownership	!	x	%	BT supply?
I.	**Recreation**				
J.	**Health**				
	J1: LL-TI of 45-pensionable age group	!	s	%	next Census
	J2: standardised mortality rate of under 75s	?	x	$	OPCS supply?
	J3: synthesis of health risk factors	?	s	#	next Census

Key

R :readiness:
- ! = appears to be fully specified, robust and relevant
- ? = robustness is still to be demonstrated (eg. via sensitivity testing of parameters in the spatial processing required)

S :source:
- s = small area Census 100% statistics
- t = TEN PERCENT small area statistics
- w = Census special workplace statistics
- x = exclusively nonCensus data

T :treatment:
- % = simple ratio
- # = different tabulations combined via a 'weighting' method
- @ = spatial average
- $ = locally interpolated wide area data

32

5 Synthesis: alternative methods for combining indicators

5.1 Whichever indicators were proposed, once it was clear that there would be more than one or two then the next question becomes how they can be combined into a single index of deprivation. This challenge in turn raises the possibility of 'weighting' the indicators according to their relative importance. In this Section, three types of synthesis methods are considered in turn. First, a number of straightforward approaches are considered, leading on to an assessment of methods which concentrated on generating 'weighting' schema for the set of indicators. The third category is a set of statistical methods which are becoming increasingly familiar.
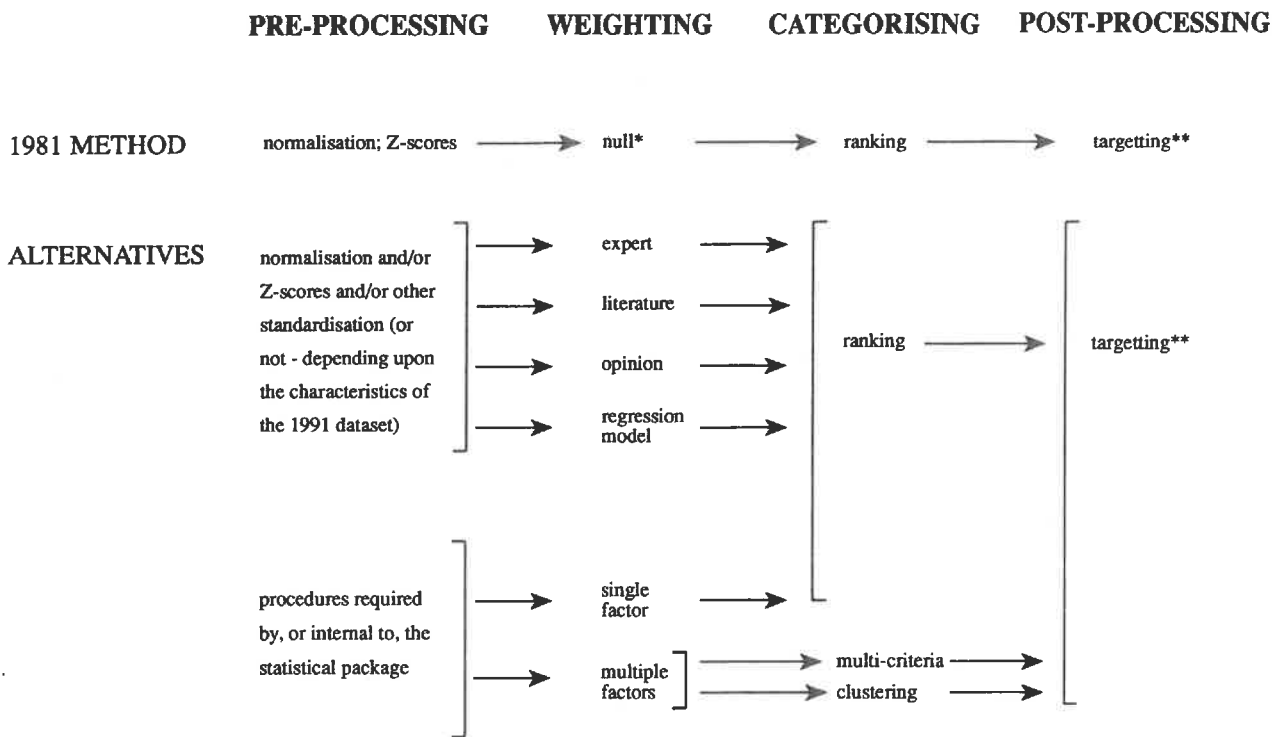
Basic alternatives

5.2 One approach which is here rejected at the outset would be derived from the 'top down' by focussing on the **Issues**. Stress has been placed many times on the assertion that the selection of indicators is too constrained by data availability for them to be considered to be perfectly representing the **Issues** involved in the concept of deprivation. Thus it would be entirely misconceived to assert that the indicators can be combined in a simplistic way which assumes that each **Issue** is accounted for neatly by one or more of the indicators which can be operationalised.

5.3 As mentioned at certain points in Section 4, a related question is whether a particular indicator in fact represents one Issue rather more than another. The most telling point, however, is that the necessary validation analysis will include preliminary correlation analyses which may cast doubt on any argument based purely on interpretation of the 'meaning' of an indicator. For example, two indicators which are each the only representative of two quite distinct **Issues** (which were thought not to be strongly inter-related) may be very closely intercorrelated. If this close association is not intuitively reasonable then entering both in the index would not appropriately represent both **Issues** in the index.

5.4 **Figure 5.1** illustrates the range of alternative approaches for combining indicators – all these methods are strictly empirical when compared to the 'top down' approach discussed above. The first alternative in **Figure 5.1** is the basic Z-scores method which was used for the 1981 analysis; the others are all somewhat more complex. Their empirical nature limits the extent to which they can be discussed conceptually, so their strengths and weaknesses need to be clarified by the validation analyses, in order to demonstrate the effect of different methods by examining the overall index which each produces. A preliminary, experimental, analysis is included in Section 6; the remainder of this Section reviews various methods' technical strengths and weaknesses.

Figure 5.1 Alternatives for the targetting 'Index' method

	PRE-PROCESSING	WEIGHTING	CATEGORISING	POST-PROCESSING
1981 METHOD	normalisation; Z-scores →	null* →	ranking →	targetting**
ALTERNATIVES	normalisation and/or Z-scores and/or other standardisation (or not - depending upon the characteristics of the 1991 dataset)	expert → literature → opinion → regression model →	ranking →	targetting**
	procedures required by, or internal to, the statistical package	single factor → multiple factors	multi-criteria → clustering →	

* In 1981 the unemployment data was double weighted to compensate for the social structure data being too late to be included.

** In 1981 this was by the proportion of the District's population in the 'top' 10% of the EDs. .

Weighting methods

5.5 The 'default' Z-scores method can be seen as a preference for applying *null* weights to the selected indicators. Champion & Green (1990) provide another example, and the comparison of their index with that of the Dept. of Environment (1983) illustrates the point that decision on how many measures of, say, unemployment are included is in effect a 'higher order' form of weighting. The apparent benefit of simplicity from this approach is also clearly a disadvantage, in that it assumes all indicators are of equal importance – regardless of the concept involved, the nature of the data available, or the objectives of any specific policy initiatives for which the ranking is needed.

5.6 The second alternative method seeks to obtain the assessment and opinions of *expert* policy makers in the specific application field. The "underprivileged area" study of Jarman (1984) is a classic example of this approach in the deprivation literature. This method has the advantage of integrating practical experience into the analysis. However, it is difficult to decide who are the experts and how to derive the precise weightings from their judgements. Of course, the results of this approach may also be open to criticism of involving personal values, vested interests and bias.

5.7 As an alternative to relying on policy experts, the weighting values can be abstracted from the *literature* by reference to a respected academic study (or studies). For instance, the points value of deprivation indicators in Townsend (1987) was based on a previous study's survey of people's expectations. However, it is unlikely

that there will be a pre-existing study which covers exactly the **Issues** identified in this study. Moreover, these weightings would need to be expressed in a set of numerical values, with one such value for each indicator generated in this study.

5.8 A public *opinion* survey on the relative importance of various deprivation issues (for example, the Rogerson et al (1989) 'Quality of Life' Study) may provide an objective measure of the public's overall views. Once again, however, it is very unlikely that such weightings obtainable 'off the shelf' from an earlier study can be matched onto the indicators generated by this or any other study that has been undertaken for a different purpose. To commission an opinion survey specially for the index to be devised following this study may not be a practical option.

5.9 There are several statistical *modelling* methods which can summarise the 'importance' of various indicators, especially if there is one single all-important measure against which they can be modelled (eg. Coombes & Raybould, 1989). This all-important indicator may perhaps only be available for one region, or for a less detailed spatial breakdown than EDs. It can then be modelled at the ED scale: that is to say, it is 'predicted' from the variables at the ED level. The biggest problem of this method is finding a valid single variable to represent disadvantage in a suitably rounded way.

5.10 This study aims to provide a general purpose assessment of deprivation. As a result, it is not plausible to derive weightings from *experts* because they tend to be focussed on particular policy concerns or programmes (eg. Jarman (1984) evaluated the factors influencing general practitioner workloads – and not the other aspects of health deprivation for which his index has been used subsequently). A more attractive option is to derive weightings from a widely respected study in the *literature* – but unfortunately there does not seem to be a consensus emerging around any past research conclusions which are embodied in a series of weightings. The same disadvantage applies to the public *opinion* weighting scheme, although the Department could consider commissioning a poll which is tailored to the finally agreed list of indicators.

5.11 The *modelling* option is also faced with the need for additional information. There is an intriguing possibility that at some further date this approach could be used to compensate for the lack of adequate income data in the local statistical sources. Survey data which covered a sufficiently large sample of people could be 'tagged' with the indicator data proposed here – via information on each individual's location in terms of EDs. The subsequent analysis might then predict the observed poverty in terms of the local deprivation indicators. The outcome would be a model, in the form of 'weightings' for each deprivation indicator, so that the level of poverty in each ED could be estimated. Once again, this would probably require a survey to be commissioned by the Department – it would also be 'reductionist' in that it would devalue non-material aspects of deprivation. An alternative version, operationalisable in two or three years, could use the microdata of the 1991 Census Sample of Anonymised Records to generate a better understanding of the way different aspects of deprivation intersect as forms of multiple deprivation (cf. Knox, 1985). The cases in the Sample would then need to be tagged (confidentially within the Census Office) and an analysis then undertaken to find which ED-level variables

predict the distribution of individuals suffering multiple deprivation. This option is too far in the future to meet the present needs, but initial discussions with OPCS about its feasibility could be worthwhile.

Statistical methods

5.12 In the absence of any of the additional information needed for the weighting system, the remaining options focus on purely empirical treatment of the indicators themselves. In short, the correlation analysis with the 1991 data has a crucial role to play. This should be extended into experimentation with, for example, factor analysis to identify whether a simple multivariate index can be derived from the indicators. All these methodologies have the disadvantage of being less easily understood, due to their higher level of statistical complexity than Z-scores. The latter technique, which is really the default method here due to its past use for the 1981 data, has already been briefly discussed in terms of being a *null* weighting approach. The remainder of this Section will take the statistical methods in turn, identifying their **strengths** and **weaknesses** after briefly describing how they are operated. To provide the default option against which they are to be compared, the Z-scores method is first described and evaluated in this way.

Z-scores

5.13 The distributions of the raw data for each indicator are checked and those which show a skewed distribution have to go through a normalisation procedure. Each variable is then transformed into a standard form so that it has a mean value equal to zero and a standard deviation equal to one. The standard scores for each indicator are then added (or subtracted, if the indicator identifies deprived areas with negative values) to produce a composite score of the relative level of deprivation. This method has four important **strengths**:

- it is a very simple and transparent method that can be easily understood;

- it will produce compatible results with 1981 if the same indicators are used, and so may provide stability in the selection of priority areas;

- the Z-scores for each area allow for targetting by ranking;

- the Z-scores can be calculated at different spatial levels from ED to District.

5.14 However, these advantages are counterbalanced by three important **weaknesses**:

- it oversimplifies the data by ignoring those relationships between the indicators which may reflect multiple deprivation;

- it is less appropriate for handling a large number of indicators unless some form of weighting is introduced, and the preceding discussion found no satisfactory source of weightings;

- there is no specific treatment of indicators which are highly inter-correlated, leading to the danger of 'double counting' (ie. indirect weighting).

Factor analysis

5.15 Factor analysis is used to identify a relatively small number of factors ('super variables') which can be used to represent relationships among sets of many variables. All the raw data will be automatically standardised in the statistical procedures before the factors are extracted. The technique seeks to summarise as

much as possible of the variance in the dataset within the minimum number of factors. Various outputs can be obtained from this technique:

(a) a **single factor** solution – identifies the nearest possible solution to an all-embracing factor which explains as much of the variance in the dataset as possible;

(b) the **first factor** solution – the first factor, in terms of its explanatory power is taken from a set of factors which have been selected because they collectively explain most of the variance in the dataset;

(c) **multi-factor** solution – more than one factor is chosen from the same set as is used in (b); these factors represent different dimensions of the dataset and can then be input to a multi-criteria analysis.

5.16 All these variants have a number of **strengths**:

- the factor(s) help to clarify the general concept – eg. 'multiple deprivation' – on the basis of the empirical links within a set of indicators;

- factor scores can be obtained for each areal unit at different spatial scales (eg. ED, ward) so that the value for each area on that factor can be used for ranking;

- the technique provides an automatic statistical weighting of each variable on the factors;

- it can deal directly with inter-correlations within the dataset by examining the correlation matrix produced in the statistical procedures.

5.17 Equally, factor analyses can be seen to have some **disadvantages**:

- the application of factor analysis involves critical decisions, including which statistical options should be used in the statistical procedures, and which factor(s) should be used for ranking;

- the single factor solution is less likely to be suitable for the analysis if a large number of indicators are included in the dataset – it works best if it succeeds in explaining a large percentage of the variance in the whole dataset.

Multi-criteria analysis

5.18 As stated earlier, the results from a multiple factor analysis can provide the basis for a Multi-Criteria Analysis. The factor scores for the chosen factors for each ED are standardised into Z-scores. If an ED's Z-scores exceed a threshold value on a set number of factors, then that ED is deemed to be multiply deprived. The method shares some of the **strengths** of Factor Analysis:

- as with factor analysis, weightings of the indicators are automatic from the factor loadings obtained from the statistical analysis;

- this method can distinguish which areas are deprived on which particular factors.

5.19 However, the method has two key **weaknesses**:

- the operation of this method requires most explanation and clarification;

● no simple ranking can be calculated for the individual EDs, so the targetting cannot be by identifying the "top 10%" most deprived EDs, (although it is possible to rank Districts on the basis of the number of EDs falling into the deprived categories).

Cluster analysis

5.20 Cluster Analysis is a statistical technique which aims to classify areas into relatively homogeneous groups on the basis of their similar characteristics. The clustering procedures for EDs will be more robust if they are based on the factors extracted from factor analysis. The characteristics of each cluster can be identified from the descriptive statistics of each variable for each cluster. The method has important **strengths**:

● it can provide a very parsimonious solution by identifying the most deprived areas in one cluster;

● it takes into account the different dimensions of deprivation within the classification process;

● the weightings of the indicators is automatically done in the statistical procedures.

5.21 Equally, there are notable **disadvantages:**

● cluster analysis requires operational decisions throughout the whole statistical procedures to attain the 'right' answer;

● since it involves two multivariate techniques, much explanation and clarification will be required to achieve general understanding;

● no rankings can be obtained at the ED level.

5.22 The recommendations will reflect the emphasis between simplicity, statistical robustness and flexibility (eg. rankings at both the ED or District level). However, policy applications always place great emphasis upon the defensibility of the method. A set of options for obtaining weightings have already been rejected on the grounds that they are all either impractical or arbitrary and potentially contentious. The null weighting implied by Z-scores is really no less arbitrary. Multi-criteria analysis requires operational decisions which are equally arbitrary. The decision on how many, and which clusters are to be deemed 'deprived areas' can be seen as contentious, although cluster systems are now widely used and can usually be shown to be statistically robust.

5.23 Recent methodological reviews by Bartholomew (1988) and Bell (1990) both assessed factor analysis as the most robust approach to combining indicators. However there is an important preliminary role for validation analyses of the assembled dataset to identify which variables are inter-correlated (see Section 6). To some degree the nature of the dataset may also prove to be more readily analysed in one way than another, but an initial preference can be stated here. A single factor solution looks likely to be the ideal, but if it is unable to capture enough of the important multi-dimensionality in the dataset then a cluster analysis should be assessed, with Z-scores remaining as the final 'default' option.

6 Verification and experimental analyses

6.1 The interim conclusions of Section 5 laid the greatest stress on the need for an empirical assessment of the 1991 dataset once it can be compiled. In particular, it is not possible to anticipate how far the recommended indicators will tend to duplicate each other. The most appropriate way of combining the indicators will be one which draws out the 'major dimensions' of that dataset in as robust a way as possible. The conclusion here was that there are no suitable weighting schemes that can be derived from existing sources, and it is assumed that the Department does not plan to commission a public opinion survey or create some policy-based weightings. Consequently, the empirical experimentation with the 1991 dataset can be limited to the statistical methods (eg. factor analysis) plus the approach used in 1981. This experimentation will be part of the verification analyses which are outlined next; the remainder of this Section then presents the results from a preliminary experimentation with 1981 data.

Verification with 1991 data

6.2 The first motivation for verification analyses stems from the proposed adoption of some entirely new indicators. It will be necessary to examine these empirically to provide reassurance that they are measuring the **Issue** concerned in the way which was intended. Some of this analysis will focus on simple rankings of the data as it comes available, in order to show that the indicator does not provide counter-intuitive assessments of which areas have the highest and lowest values for this aspect of deprivation.

6.3 This first stage of the verification analyses is particularly necessary for the GIS-based measures. In most GIS operations there are critical parameters to be set and operational decisions to be made (Martin, 1991). The most important GIS technique needed here is that which produces for each ED a *spatial average* from the values of adjacent areas. These input values may be at the ED level, providing a "generalisation" of the data where there is uncertainty about data accuracy for individual EDs – or they may be for larger data reporting areas, thereby achieving an "interpolation" from those fewer input values to the larger number of EDs (Coombes & Raybould, 1990). The extent to which spatial variation in the input data is smoothed by the spatial average can be controlled by parameter setting in the GIS. In general, the level of smoothing should be determined by an understanding of the local variation of that aspect of deprivation, but it will also be necessary to examine the data generated and to fine tune the parameters by assessing experimental results.

6.4 The second motivation is the need to appreciate the inter-relationships between the indicators. As has been stressed several times, the availability of data inevitably results in the proposed indicators falling some way short of perfectly representing

their particular **Issue** of deprivation. Obtaining most indicators from the Census can lead to implicit bias in the data, so that measures which are intended to represent different **Issues** are in fact highly correlated. It is essential to carry out correlation analyses, although it will not *always* be appropriate to remove one indicator where two are found to be highly correlated. The decision will depend on whether the two indicators were intended to be representing very distinct patterns, and on which synthesis method is to be used.

6.5 The question of which method is to be used to synthesise the indicators into a single index is the third motivation for the verification analysis. It is not necessary here to rehearse the options involved in this question, but the remainder of this Section provides a preliminary, experimental, analysis of 1981 data using various statistical methods.

Preliminary experimentation with 1981 data

6.6 To provide an empirical background to the methodological experiments in the verification analysis, a preliminary study has been carried out on 1981 data. As described in the **Appendix**, a multivariate dataset was created at the ED level to contain most of the properties which the recommended 1991 dataset is likely to possess (eg. including some variables which depended on GIS techniques). A whole series of large scale analyses were undertaken on this very substantial dataset: the **Appendix** includes the details of the statistical operations which were carried out. Considerable effort was devoted to check that broad patterns observed were not specific to one particular analysis, but were reproduced by other variants of the same type of approach.

6.7 All the approaches pursued involved a standardisation of the variables which is analogous to creating the Z-scores which provided the main feature of the earlier 1981 analysis (Dept. of Environment, 1983). In the more complex analyses, this procedure is a preliminary one within a multi-step process (such as cluster analysis). The major distinction is between the basic method and the others considered, all of which involve a second step of creating multi-variate factors or principal components from the dataset. These factors have the advantage of identifying the major statistical dimensions in the data – and removing problems of duplication between variables – but they then require interpretation.

6.8 For the basic approach, the **Z-scores** are simply summed across all variables – given the earlier rejection of any of the available weighting schema. This then provides a 'score' for each ED; a similar score can be computed for a multivariate factor. Two variants of this latter approach are considered here – one is from an analysis that sought a **single factor** to summarise as much variance in the dataset itself, the other was the **first factor** from an analysis which recognised nine other distinct dimensions in the dataset (as detailed in the Appendix). In each case, the single factor of interest can be seen as an equation that effectively weights the various indicators to different degrees; this equation can then be applied to each ED's data to compute its overall score.

6.9 The other two approaches are quite distinct, in that they do not generate scores for EDs. They do, however, classify which EDs are deprived, so that wider areas such as Districts can be ranked by the percentage of their population which is in the deprived EDs. First, a simple **multi-criteria analysis** has been applied to the

first three factors from the 10 factor analysis underlying the previous approach. Each ED's value on each of these 3 factors is then translated into a Z-score. The examplar analysis here focussed on EDs with values of over 0.5 (ie. above the mean by more than half a standard deviation on that factor). EDs were then deemed to be deprived if they 'qualified' on two out of the three factors in this way. It is notable that although this approach may appear unfamiliar, it finds echoes in the early studies by Holterman (1975) and Berthoud (1983).

6.10 The final approach is the more familiar one of **cluster analysis**, which in this context used to be categorised as social area analysis (eg Cullingford & Openshaw, 1982). This method also carries out standardisation and factor analysis, as part of a suite of routines leading to groupings of EDs with high levels of multivariate similarities. Cluster analyses on this scale – nearly 100,000 EDs – are still major exercises, especially since many have to be undertaken to select the most appropriate groupings (because these vary with the number of clusters defined). The selected analysis here had 7 clusters, of which 2 were identified as focussing on deprived areas (see the **Appendix** for further details).
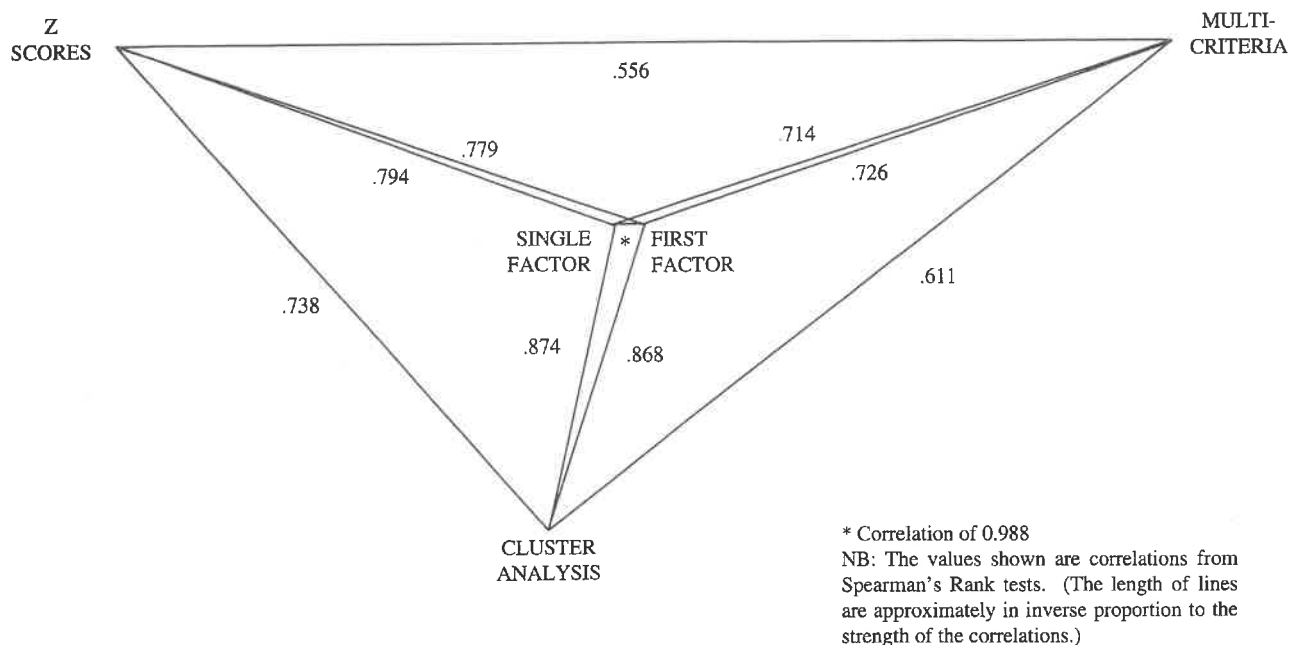
6.11 These last two approaches have identified lists of EDs which are deprived, rather than scores for all EDs which can be ranked. At the local authority scale which is of primary interest for policy purposes, these results can be converted into values for each local authority, which can then be ranked. It would be possible to carry out a similar step for the other analyses – just as the original 1981 analysis used the Z-scores to calculate the proportion of each local authority's EDs that were in the top 10% of the ranking. However this step can distort and/or lose much of the original information and is not necessary for this experimental analysis. In effect, all that is needed here are rankings of local authorities by their actual scores (in the first three methods), or by the percentage of their EDs which are categorised as deprived (by the last two approaches).

6.12 Spearman rank correlation analysis allows the similarity of the results from the five approaches to be assessed. The results are portrayed graphically in **Figure 6.1**, which shows that the most similar approaches are the two factor analyses, which are virtually identical. They are also in effect the modal form of analysis: all other approaches are more similar to the factor analyses than to any other set of results. Remarkably, cluster analysis is the other approach which is closest to these two – despite it having generated categorical data which then had to be simplified as the proportion of each District's EDs which are in the deprived clusters. The Z-scores approach yields the next most similar set of results, with the multi-criteria analysis giving the most distinct results. This last finding is not surprising because this approach has least technical similarities with the others.

6.13 It should also be remembered that there is no objective assessment here of the 'best' mapping of deprivation among the five alternatives. It is also worth noting that all five methods found Districts both from Inner London and from northern metropolitan Counties among their ten most deprived areas. The results appear to be sufficiently similar and intuitively reasonable to provide early guidance on the possible findings from the 1991 validation analyses. The methods also give different results to the extent which might have been expected from their particular strengths and weaknesses. Of course, the use of the 1981 data makes the detailed rankings of only incidental interest here.

Figure 6.1 Similarity between analyses

Figure 6.1 Similarity between analyses

Z SCORES

MULTI-CRITERIA

.556

.779

.714

.794

.726

SINGLE FACTOR

* FIRST FACTOR

.611

.738

.874

.868

CLUSTER ANALYSIS

* Correlation of 0.988
NB: The values shown are correlations from Spearman's Rank tests. (The length of lines are approximately in inverse proportion to the strength of the correlations.)

6.14 For the purposes of this study, these results provide little reassurance in the face of past criticism of a Z-scores ranking approach. The major methodological difference between it and the other methods here is its retention of each indicator as a distinct and equally significant element of the analysis. This may appear to be a strength in the approach - but this depends entirely on the indicators all being very well defined, being proven to be statistically distinct from each other, and being known to be measuring equally important aspects of deprivation. This seems a severe set of assumptions at this stage. In the experimental analysis, as much detail as possible was retained by calculating each local authority's mean score across all its EDs.

6.15 If the results from the different approaches can be seen as representing the distilled 'collective wisdom' from the dataset, then it is discouraging that the results of the Z-scores approach were something of an outlier in **Figure 6.1**. The conclusions here, then, are that a basic Z-score analysis cannot be assumed to provide results which are virtually interchangeable with those from more complex approaches. The larger number of variables recommended here, in comparison with 1981, probably leads towards a method involving factor analysis, but the 1991 dataset will need exploratory analysis in order to identify the most robust and appropriate form of analysis for identifying areas of deprivation (Bell, 1990). However, there will be another important judgement to be made, even after selecting the most appropriate methodology: it will be necessary to stress that analyses were undertaken which showed other methods were less satisfactory, but it would be confusing and counter-productive if the results of each method were published (with each showing a different result for every area). The form and content of the presentation of the validation analysis results will be critical in strengthening the confidence in the 1991 index.

7 Monitoring: handling change over time

7.1 In the discussion over competing priorities for this study (Section 3) the preference for a dynamic aspect to the analysis was identified, but was then found to conflict strongly with the other objectives. As a result, the concern with change over time has largely been set aside until this point; it is now possible to assess how far the approach recommended so far can be enhanced with a dynamic perspective. The main priority is for updating after 1991, but the first question is the relationship of the new analysis to the 1981-based results. The default option was stated as re-running the 1981 method on 1991 data: the first issue considered here is whether there are better alternatives.

Looking at change since 1981

7.2 There is an alternative strategy for gaining a view of the trends which led up to the distribution of deprivation which will be observable with the 1991 data. This strategy would seek 1980s data for the indicators which are proposed here, instead of seeking 1991 data for the indicators in the original Z-score analysis of 1981 data. Of course, many problems would be encountered in trying to implement this strategy. Some of the sources relied upon here for 1991 have only been publishing data for a few years (eg. the insurance companies for the crime risk data) while others are still dependent on a data release agreement (eg. on telephone ownership) which is highly unlikely to extend to providing data for several years ago. Part of the attraction of several of the Census indicators was that they are much improved versions of their 1981 equivalents (notably ethnicity and LL-TI), so backdated versions of these indicators too will be flawed.

7.3 The nature of the analysis requires that a rigorous study of trends would need to use data that was very closely matched in 1981 and 1991, so the conclusion must be that this backdating strategy is not feasible. Trends for 1981-91 can thus be more accurately (and inexpensively) calculated using 1991 data with the 1981 indicator specification. Of course, ED boundaries are not static, so the results will not be comparable at the finest level of detail, except for the (non-random) sample of EDs which have not changed. It should be possible to approximate 1981 wards from groupings of 1991 EDs, so a nationwide comparison on stable areas will be feasible at this level. However, the aim is not to discover whether each neighbourhood is x% more or less deprived than in 1981. The greater concern is with the relativity between areas, and the areas of policy concern are more often at the District level. It will be perfectly possible to compare the 1981 and 1991 proportions of each District's population "in the worst 10% of all EDs" (based on separate analyses of the rather different 1981 and 1991 EDs). The remaining

problem is that there will be two somewhat different portrayals of deprivation in 1991, which could lead to some confusion unless the 1991 data for the 1981 indicators is very carefully presented.

Updating in the 1990s

7.4 The analysis based around the 1991 Census data cannot be 'the last word' on deprivation, even if its conclusions succeed in attracting a high level of consensus. This is not least because much of the data will already be nearly 3 years out-of-date by the time it is published. If this problem implies the need for publication of the results at the earliest opportunity, there could still be 1981 data included in that analysis: the job accessibility measure [E5] needs commuting data that will not be available until late 1994, and the availability of the 10% SAS data which is used for several indicators (see **Table 4.1**) is currently under review (OPCS & GRO(S), 1992b). At present, the assumption is that this dilemma will be faced by putting priority upon the earliest possible publication – so there would be an inherent need to update that analysis as soon as the 10% sample 1991 data is published.

7.5 Updating at very frequent intervals would probably create an inappropriate impression of instability to the analysis. However, waiting until the next Census would be to rely too heavily upon the continuity of the patterns observed (which is due to the deep-seated nature of the problems). It might be adequate to update only twice per decade if it was known that there would be a 1996 Census – but in fact there is a current question mark over *any* Census in the future (OPCS, GRO(S) & CO(NI), 1992). In this context of uncertainty, it is only really feasible here to consider the possibility of updating the proposed indicators from nonCensus sources by the mid-1990s.

7.6 Of the 21 recommended indicators, only 4 [**A1, B1, E4, and J2**] are from nonCensus sources which provide regular updates – in fact, all are at least annual. One or two of these sources would generate more robust indicators if several years' data was combined together, but this still allows for an annual rolling forward of the data if necessary. Two sources [**for G3 and H2**] still depend on negotiations which *might* lead to annually updated data becoming accessible, but it may be more realistic to aim for access on alternate years. Frequent updating would certainly be of interest for internal monitoring, but its publication could lead to pressure for repeated re-targetting of policies (which would create more uncertainty than it would resolve by appearing to be more up-to-date).

7.7 With the aim of at least one update by the mid-1990s, then, the question is how far proxy data might be available for the bulk of the Census-based indicators. The private sector is increasingly active in estimating local level data, but in almost all cases they use standardised models (eg. of residential turnover) which will thus ignore the possibility of different areas having different trends – the very possibility which is of interest here. The 1991 male unemployment measure [E3] was designed to be from Census data in order to obtain data at the ED level, but the Dept. of Employment's alternative source is updated monthly. The problem here is that the updated denominators needed for the percentage rate calculation are not made available at the ward level (as is the unemployment numerator) but for the wider Travel-to-Work Areas. Unofficial alternatives are available but have not been accepted by Government statisticians. One technique could be to use GIS techniques

on the official data, for 1991 and later, to generate a 'trend surface' across the country from which each ED's probable trend could be estimated. This might be adjusted according to the ED's demographic structure, and the 1991 Census denominator then projected forward accordingly. An annual average from the monthly unemployment count could then be used to estimate a local male unemployment rate for the mid-1990s. It would be important to collaborate closely with the Dept. of Employment in commissioning research to create a viable method for estimating small area workforce data.

7.8 This example serves to show that it is not a trivial matter to estimate ED-level data with any degree of conviction. Of course, more profound difficulties arise where there is simply no available source on which to base the estimates. For example, moving to the indicator of differential ethnic minority unemployment [A2] leads to a complete gap in nonCensus data. Nor is it possible to piece together a cross-section of the 21 indicators so that there is one which is updatable for each **Issue**, because all 3 Housing indicators are from the Census and there is no alternative source which provides consistent localised information on any of these concerns (amenities, mismatch, overcrowding). This difficulty then raises the question of whether it is worthwhile investing a great deal of resources in collating data to provide some updating of, effectively, a non-random selection of indicators. For example, local education service data might be obtainable to allow better monitoring of the local proportion of 16 and 17 year olds who 'stay on' [D2], but substantial resources would be needed to ensure that the data from every area was sufficiently detailed in its spatial coding to provide valuable information here.

7.9 The more general point raised here is whether there are a substantial number of potentially valuable indicators which are available at the District level but cannot be made available for more detailed areas. It is certainly hard to identify any of the District level data in *Regional Trends* (Central Statistical Office 1991) as important indicators of deprivation which have had to be missed in Section 4's recommendations for ED level data. For example, the relevant information on **Education** (eg. truancy or Careers Service records of school-leaver unemployment) is not routinely broken down to the District level in Shire Counties. A small number of Dept. of Environment datasets could be of interest – the derelict land and house conditions surveys in particular - but these have already been discussed as potential sources for neighbourhood level data (given fully disaggregated data, plus research on appropriate techniques to generate estimates at the finer scale). The other potentially interesting source is OPCS, and in particular the cause-specific mortality data which could be analysed to identify "avoidable deaths" in each area (cf. Coombes et al, 1992). This data is probably not processable to a finer level of detail and so is perhaps the one clear case of a deprivation measure which could be added to an index operated purely at the District level. Of course, there are other policy related District level datasets – notably the formation and closure of firms, and estimates of net migration – but these are not really aspects of deprivation.

7.10 The general conclusion appears to be fairly gloomy for the prospects of updating the ED level index. However, the necessary experimentation with 1991 data might just provide some assistance. It is possible that the indicators for which updating is scarcely feasible will prove to be very closely correlated with ones which can be updated. Thus the eventual choice of an index method could be in

part guided by the benefits of updatability – choosing not to use a Census-specific indicator when an indicator from an annually published source appears to be representing much the same pattern across the country. Of course, the fact that all the proposed indicators for some **Issues** (eg. those on **Housing**) are from the Census makes it unlikely that these indicators can be proxied from nonCensus sources. The best which can be realistically expected might be that a secondary index could be developed, using only updatable indicators, to provide a 'forward looking' 1991 analysis which was as close as possible to the definitive all-indicator 1991 targetting analysis.

8 Conclusion

8.1 A methodological study cannot generate conclusions which, for example, reveal who is suffering deprivation, or where policy should be targetted. This study has sought answers to five broad questions concerning the way that targetting should be undertaken with 1991 data: the following paragraphs summarise these methodological discussions in turn.

8.2 The first question addressed the *concept of deprivation* itself. Reviewing the diverse literature related to the concept proved of limited help in guiding a targetting analysis. There are some 'high level' principles on which many authorities are agreed – although the literature also includes substantial diversity of approaches, not to mention several cases that are simply confused. In particular, there is often a detachment between the conceptual discussion and the subsequent statistical analysis (with perhaps a casual acceptance that data limitations enforce an approach which is not in keeping with the conceptual discussion). In this study, the fact that the targetting analysis would be using spatial data is recognised at the outset, so that the concept of deprivation is framed in terms which do not depend on a subjective assessment, and do not imply a severe distinction between prevalence and risk. This approach is allied to an emphasis on the interdependence of different aspects of deprivation. The literature does not provide a consensus on the identity of these aspects or **Issues** (eg. **Income, Social Conditions**) so a set of ten **Issues** were devised here, using as a starting point the UN's Level of Living schema.

8.3 The second question was the intermediate one of the most appropriate *approach to the targetting analysis*. The spatial element of the approach was emphasised in the concept – ie. the targetting of areas of deprivation – but it left unclear the 'units of observation' for that analysis. The ideal unit would be a neighbourhood, but there is no readily available set of boundaries which represent such units across the country. For the short term, the only viable option is the ED: this is too small for several of the **Issues** under examination here, but GIS techniques can be used to recognise the context for each ED (and also make its statistical profile more robust). The approach adopted here has been to place the greatest emphasis upon robustness, with a corresponding reduction in the importance of other objectives for the study.

8.4 One particular implication of this approach is found in the answers to the third question, which focussed on the *selection of indicators*. Two of the **Issues** were left without any indicator that was considered sufficiently robust at the very localised level of the intended analysis. The association, to one **Issue** rather than another, of several proposed indicators remains open to debate. The basic approach here justifies this relaxed attitude to 'labelling' the indicators: all indicators are known to be imperfect, and most of them offer rather indirect perspectives on just one or other

aspect of an **Issue**. The crucial point here is the rejection of any claim to have very tightly identified any indicator – culled from a very general source such as the Census – with a theoretical discussion of the concept of deprivation. A more realistic approach, especially for spatial analysis, stresses that there are interconnected syndromes of deprivation and that the task is to seek indicators which between them can reflect most of the main dimensions of these syndromes. From this standpoint, the 21 proposed indicators are thought to provide a relatively strong average of the ten **Issues** of deprivation.

8.5 The corollary of the relaxed approach to assigning indicators to **Issues** is that the Issues do not provide a framework which answers the question of how to combine the indicators in a *synthesis as an index*. In fact the approach here once again is to be very pragmatic, in part because of the known limitations of the available data. The major alternative approaches are discussed, and several statistical methods are identified as options for the 1991 analysis. An experimental analysis is undertaken with 1981 data, demonstrating that different methods generate appreciably different results and that the Z-scores approach is neither the most similar nor the most different to the others considered. The main recommendation is that there must be an evaluation of the 1991 data when it is collated; it is not possible to firmly select the most appropriate techniques in advance of appreciating the dataset's characteristics. Whether factor analysis or cluster analysis is considered the leading candidate is largely dependent upon whether a ranking is required at the ED level.

8.6 The technical discussion of methods for creating an index led to the rejection of several options, but numerous others remained to be assessed as part of a *validation analysis*. The initial motivation for undertaking some preliminary analyses with the 1991 data is that several indicators use sources and/or methods which are new. The need to evaluate the data, and perhaps to adjust the indicator's detailed specification, is particularly important for any measure which depends upon the use of GIS. The second motivation for the validation analysis is the critical need to identify the statistical relationships between indicators through correlation analysis. This leads into the final stage of selecting the most appropriate form of synthetic index: here again experimentation will be important. From the preliminary experimental analysis carried out in this study, the most promising approach appears to be based on factor analysis, although cluster analysis would be a possible alternative if a single factor cannot represent all the main dimensions of the dataset. The default alternative remains the Z-scores method.

8.7 The sixth and final question concerned the feasibility of setting the 1991 targetting within an *analysis of trends*. First, the change in the areas targetted, when compared to those from the 1981 analysis, needs to be understood. The recommendation is to re-apply the 1981 indicators and method on 1991 data, so as to both assess the change over the decade and also appreciate the difference in 1991 between the two methods' results. As for change after 1991, the prognosis is not bright. A number of components of an updating strategy were set out, but the maximum achievable result remains limited – at least before any data from the possible 1996 Census. The major response to this problem could be to include in the evaluation of the 1991 dataset an exercise that is limited to those indicators which are updatable. If a 1991 index with this subset of indicators can be devised

to quite closely mimic the full index, then this could provide the basis for an updating approach. To claim to update the full index by just updating a small number of its indicators is probably not statistically acceptable.

8.8 Running right through the study has been an emphasis upon the spatial aspect of targetting: this spatial aspect has often been dealt with explicitly by recommending *the use of GIS* techniques in the indicators' development (see **Annex**). Without doubt there would also be very substantial benefits from collating the 1991 dataset within a GIS: this would maximise the immediacy that comes from mapping, it would facilitate the linkage to higher level areas like Districts, and it would stimulate experiments with methods which, for example, identify which EDs with certain characteristics are located close together In short, a GIS provides a valuable 'toolbox' for a large spatial analysis exercise such as targetting; this study has shown that for the targetting analysis to succeed in identifying areas of deprivation it will be necessary to bring together many datasets with a range of software – plus expertise which extends from spatial analysis researchers to specialists in contemporary social problems.

Annex: detailed specification of the proposed indicators

A. Social Conditions

A1 Crime rate

Definition: The cost of insuring 1,000 pounds of property contents is used as a proxy for crime risk exposure.

Source(s): The area premium charge rates are listed in the underwriting guides of insurance companies for each postcode district in the country. It is important to note that some companies do not provide a full list of their charge rates as special applications are required to insure home contents in very high risk areas. However, there does not seem to be a shortage of companies which do provide full information.

Operation: The average home content insurance costs (of several companies) levied in each postcode district is calculated. For each ED, the value will be a spatially weighted mean of values for areas within 5 kms of the ED (or the nearest postcode district, if none is within 5 kms).

A2 Differential ethnic minority unemployment

Definition: Comparing the differences in unemployment levels between the 'non-white' and the 'white' population.

Source(s): The statistics will be available in the 1991 Census SAS Table 9 (Economic position and ethnic group).

Operation: Unemployment rates for the non-white and the white population groups are calculated for each ED. The differential between the ethnic unemployment rate and the white unemployment rate could then be obtained for each ED. Subsequently, the value for each ED is calculated as a 'spatially weighted average' from all EDs within, say, 3 kms of that ED's centroid. Any negative values (probably due to small numbers) would be set at zero.

B. Physical Environment

B1 Road traffic accidents to children

Definition: The aggregated risk of road traffic accident (RTA) to child pedestrians aged between 5-15 over the nearby area.

Source(s): The Department of Transport holds a grid referenced database "STATS 19" of all reported accidents and includes the age of the casualty.

Operation: RTA grid referenced data are used to calculate the aggregate 'risk' – in terms of the cumulative value (weighted by distance using GIS) of RTAs within, say, 5 kms - for each ED.

C. Housing

C1 Housing without full amenities

Definition: Percentage of persons living in accommodation without full amenities (ie. which does not have central heating and exclusive use of a bath and WC) or in non-permanent accommodation.

Source(s): The statistics will be available in the 1991 Census SAS Table 20 (Tenure and amenities).

Operation: (all people in households – all people in permanent households with all amenities)/(all people in households) * 100

C2 Households in mismatched accommodation

Definition: Number of households with pensioner(s) or dependent child(ren) living in accommodation without central heating and/or in non-selfcontained accommodation as a percentage of all households with pensioner(s) or dependent child(ren).

Source(s): The statistics will be obtainable in the 1991 Census SAS Table 46 (Households with dependent children: housing); Table 47 (Households with pensioners: housing).

Operation: (households with dependent child(ren) without central heating and households with pensioner(s) without central heating and households with dependent child(ren) in non-selfcontained accommodation and households with pensioner(s) in non-selfcontained accommodation/(all households with pensioner(s) or dependent children) * 100

C3 Rooms shortfall from 'Standard'

Definition: Number of rooms shortfall (ie. rooms needed to meet with the standard of one person per room) as a ratio of all households.

Source(s): The statistics will be available in the 1991 Census SAS Table 22 (Rooms and households size).

Operation: number of rooms shortfall from the position where the number of persons is never more than the number of rooms in any household (eg. a household with 4 persons in 2 rooms is 2)/all households.

D. Education

D1 Low level of education (Low Skill SEGs)

Definition: Number of economically active residents in households with the household head classified into socio-economic groups 7, 10, 11 or 15 as a percentage of all residents who are economically active.

Source(s): The data will be available in the 1991 Census SAS Table 86 (SEG of households and families 10% sample).

Operation: (all economically active persons in SEGs 7, 10, 11 or 15)/(all economically active persons) * 100

D2 16 to 17 year olds not in education

Definition: Percentage of 16 & 17 year olds not in full-time education.

Source(s): The data will be available in the 1991 Census SAS Table 10 (Term-time address) and Table 37 (Young adults).

Operation: (all students aged 16 & 17)/(all persons aged 16 & 17) * 100

E. Employment

E1 Hours worked per non-dependent

Definition: Weekly working hours of those in employment per non-dependent (ie. those of working age).

Source(s): The statistics will be available in the 1991 Census SAS Table 75 (Hours worked 10% sample).

Operation: (total weekly working hours of all males and females in employment)/(all those of working age) * 100

E2 Low skill SEG unemployment rate

Definition: The proportion of economically active residents in SEGs 7, 10, 11 or 15 that are unemployed.

Source(s): The statistics will be available in the 1991 Census SAS Table 92 (SEG and economic position 10% sample).

Operation: (unemployed in SEGs 7, 10, 11 or 15)/(all economically active residents in SEGs 7, 10, 11 or 15) * 100

E3 Male unemployment

Definition: Percentage of economically active males that are unemployed.

Source(s): The statistics will be available in the 1991 Census SAS Table 8 (Economic position).

Operation: (number of males that are unemployed)/(number of economically active males) * 100

E4 Likelihood of ceasing unemployment

Definition: The likelihood of unemployment benefit claimants ceasing to be unemployed in Winter.

Source(s): The statistics are available from NOMIS as dataset PFD.

Operation: The value is calculated as the (deduced) flow of male claimants leaving the register divided by the average number of male unemployed over the six months to April 1992.

E5 Access to job opportunities

Definition: Average commuting distances of low skill groups.

Source(s): The data is in the Census SWS – but the 1981 ward data will have to be used until 1993/94 (it is not yet decided whether the 1991 data will be available for EDs).

Operation: For all residents in employment of SEGs 7, 10, 11 or 15, the median distances of commuting flows is calculated from Section A of the SWS. For each ED, the value will be a spatially weighted mean of values for areas centred within 5 kms of the ED (or the nearest ward, if none is within 5 kms).

F. Work Conditions

No proposed indicator.

G. Income and Needs

G1 'Adult Equivalence' dependency

Definition: The calculation of Adult Equivalent Scales is to provide a more refined measure of needs to account for the fact that not everybody has identical needs, and also there are economies of scale in household consumption. The adult equivalent scales adopted here are the ones recommended by the OECD, they are:

first adult in household	1.0
each other adult	0.7
each child	0.5

Source(s): The statistics required will be available in the 1991 Census SAS Table 31 (Dependent children in households) and Table 34 (Economic position of household residents).

Operation: A ratio of adult equivalent dependents (ie. total adult equivalent needs – full time equivalent employment) to adult equivalent needs of an area.

G2 Children in non/low earner households

Definition: Percentage of dependent children in households without any earners or in lone parent households with part-time employment.

Source(s): The statistics will be available in the 1991 Census SAS Table 36 (Earners and dependent children) and Table 40 (Lone parents).

Operation: (total number of children aged 0-15 in households with no adult in employment and number of children in male/female lone parent households with part-time employment only)/(all dependent children) * 100

G3 Children in households on Income Support

Definition: The proportion of children who are in households receiving Income Support.

Source(s): The statistics are being requested from the DSS (Benefit Agency) – preferably to be commissioned for analysis in Summer 1992. 1991 Census SAS Table 42 (Household composition and housing).

Operation: The recipients of Income Support are 99% fully postcoded so DSS can count the number of children in such households for each postcode unit. The data can be matched to EDs by DSS, who keep the Central Postcode Directory (although they do not yet have a version including 1991 EDs). These counts can be processed

into rates by matching with 1991 Census data on households with children (suppressing in rates over 100%). It may be appropriate to generalise the data slightly by GIS techniques in order to avoid any confidentiality problems.

H.   Communications

H1   No car & poor public transport

Definition: The availability of public transport to residents in households without a car but require some transport to work.

Source(s):   The statistics will be available in the 1991 Census SAS Table 82 (Travel to work and SEG 10% sample) and Table 86 (SEG of households and families 10% sample).

Operation: (% of all residents in households without a car)* [1- (% of all workers in non-car household that require some transport to work use public transport/100)]

H2   Telephone ownership

Definition: The proportion of all households that have a telephone.

Source(s):   BT postcoded information and the 1991 Census SAS Table 28 (Dependents in households).

Operation: Presuming that BT makes available counts of private lines by postcode units, these will be summed (using the census postcode file) to ED and compared with the numbers of private households (values over 100% will be set at 100%)

I.   Leisure

No proposed indicator.

J.   Health

J1   Limiting long term illness of aged 45 to 59/64

Definition: Percentage of males and females aged 45 to pensionable age in private households with limiting long term illness.

Source(s):   The statistics will be available in the 1991 Census SAS Table 12 (Long term illness in households) and Table 11 (Persons present).

Operation: (number of males aged 45 to 64 in private households with limiting long term illness and number of females aged 45 to 59 in private households with limiting long term illness)/(number of males aged 45-64 in private households and number of females aged 45-59 in private households) * 100

J2 Standardised mortality rate of under 75s

Definition: Age-sex standardised mortality rates of people aged under 75.

Source(s): The ward level data has to be requested from OPCS.

Operation: The age-sex standardised mortality rates (disregarding cause of death) can be calculated for those aged under 75, provided that OPCS agree to the provision of this data for wards. The data is then spatially interpolated to the ED scale, which will also help to reduce any concerns over confidentiality or statistical qualms at publishing the raw ward level data.

J3 Synthesis of health risk factors

Definition: The proportion of all household residents that are considered as vulnerable to 'health risk', they are residents in:

(1) lone parent households that are non-owneroccupied;

(2) lone parent households without a car, or

(3) households without a car and non-owneroccupied.

Source(s): The statistics will be available in the 1991 Census SAS Table 20 (Tenure and amenities) and Table 46 (Household with dependent children: housing).

Operation: (number of persons in lone parent households that are non-owneroccupied and number of persons in lone parent households without a car and number of persons in households without a car and non-owneroccupied)/ (all persons in households with residents) * 100

Bibliography

Bartholomew D J (1988) *Measuring social disadvantage and additional educational needs* Report to Department of Environment, LSE Department of Statistical & Mathematical Sciences, London

Begg I, Eversley D (1986) "Deprivation in the inner city: social indicators from the 1981 Census" in V A Hausner (ed.) *Critical issues in urban economic development: vol 1* Clarendon Press, Oxford

Bell D (1990) *Data sources for area prioritisation: Section A – review and analytical topics* DG Information Services, Edinburgh

Berthoud R (1983) "Who suffers social disadvantage?" in M Brown (ed) *The structure of disadvantage* Heinemann, London

Berthoud R, Kempson E (1992) *Credit and debt: the PSI report* PSI, London

Brown M, Madge N (1982) *Despite the welfare state* Heinemann, London

Campbell DA, Radford JMC, Burton P (1991) "Unemployment rates: an alternative to the Jarman index?" *British Medical Journal* 303, 750-755

Carley M (1990) *Housing and neighbourhood renewal: Britain's new urban challenge* PSI, London

Carstairs V, Morris R (1989) "Deprivation and mortality: an alternative to social class?" *Community Medicine* 11, 210-219

CEC Directorate-General for Regional Policy (1991) *The regions in the 1990s: fourth periodic report on the social and economic situation and development of the regions of the Community* Commission of the European Communities, Brussels

Central Advisory Committee for Education (1967) *Children and their primary schools* (Plowden Report) vol.1 HMSO, London

Central Statistical Office (1991) *Regional Trends 26* HMSO, London

Champion T, Green A (1990) *The spread of prosperity and the north-south divide* Booming Towns, Gosforth & Kenilworth

Coombes MG, Raybould SR (1989) "Developing a Local Enterprise Activity Potential (LEAP) index" *Built Environment* 14, 104-117

Coombes MG, Raybould SR (1990) "Profiling neighbourhoods: social and environmental statistics for Tyne and Wear" *NorthEast RRL Research Report 90/1* CURDS, University of Newcastle upon Tyne

Coombes MG, Raybould SR, Wong YLC (1992) *Indicators of potential for inner city regeneration* HMSO, London

Coombes MG, Raybould SR, Wong YLC (1994) *Indicators of disadvantage and the selection of areas for regeneration* Research Report 26, Scottish Homes, Edinburgh

Coulter J (1978) *Grid-square census data as a source for the study of deprivation in British conurbations* Working Paper 4, Census Research Unit, University of Durham

Cullingford D, Openshaw S (1982) "Identifying areas of rural deprivation using social area analysis" *Regional Studies* 16, 409-418

Daniel W W (1990) *The unemployed flow* PSI, London

Department of the Environment (1983) *Urban deprivation* Information Note 2, Inner Cities Directorate, Dept. of Environment, London

Duguid G, Grant R (1983) *Areas of special need in Scotland* Central Research Unit, Scottish Office, Edinburgh

Dutton P (1992) "Pensioners left in cold by village's big divide" *The Journal* 5 February, 5

Economist, The (1990) *Vital world statistics: a complete guide to the world in figures* Hutchinson, London

Evandrou M (ed.) (1991) *Improving government statistics: gaps and discontinuities in official statistics* Social Science Forum, London

Fox AJ, Goldblatt PO (1982) *Socioeconomic differentials in mortality 1971-1975* OPC5 LS1, HMSO, London

Garner C, Main BGM, Raffe D (1988) "A tale of four cities: social and spatial inequalities in the youth labour market" in D Raffe (ed) *Education and the youth labour market: schooling and scheming* Falmer, London

Goldstein H (1987) *Multi-level models in educational and social research* Griffin, London

Gray J, Jesson D, Pattie C, Sime N (1990) "The inner city dimension in Britain's youth labour market" *Employment Gazette* 98, 455-463

Green A, Owen D (1990) *Long-term unemployment: JUVOS analysis* Research Paper 72, Department of Employment, London

Hatch S, Sherrott R (1973) "Positive discrimination and distributions of deprivations" *Policy and Politics* 1, 223-240

Henley Centre (1991) *Leisure Futures* Henley Centre, London

Herbert DT, Darwood J (1992) "Crime awareness and urban neighbourhoods" in DJ Evans, NR Fyfe, DT Herbert (eds) *Crime, policing and place: essays in environmental criminology* Routledge, London

Hirschfield A (1989) "The study of poverty and deprivation in developed countries" *Urban Policy & Evaluation RRL Working Paper 2* Dept of Civic Design, University of Liverpool

Holterman S (1975) "Areas of urban deprivation in Gt. Britain: an analysis of census data" *Social Trends 6* HMSO, London

Hope T, Hough M (1988) "Area, crime and incivility: a profile from the British Crime Survey" in T Hope, M Shaw (eds.) *Communities and crime reduction* HMSO, London

Hough M, Lewis H (1988) "Counting crime and analysing risks: the British Crime Survey" in DJ Evans, DT Herbert (eds.) *The geography of crime* Routledge, London

House of Commons (1977) *Policy for the inner cities* (White Paper) Cmnd. 6845, HMSO, London

Institute of Social Studies Advisory Service (1990) *Poverty in figures: Europe in the early 1980s* Report to Eurostat, (European Communities no. CA-59-90-685-EN-C) Luxembourg

Jarman B (1984) "Underprivileged areas: validation and distribution of scores" *British Medical Journal* 289, 1587-1592

Kearns A (1990) "Housing policy, deprivation and space: the case of stress areas" *Policy & Politics* 18, 119-134

Knox PL (1985) "Disadvantaged households and areas of deprivation: microdata from the 1981 Census of Scotland" *Environment & Planning A* 17, 413-425

Lawless P (1979) *Urban deprivation and government initiative* Faber, London

Lawson S (1991) *Accidents to young pedestrians: distributions, circumstances, consequences & scope for counter-measures* AA Foundation & Birmingham City Council

Lever B, McGregor A, Paddison R (1990) *City audit: comparative European urban performance data* CURR/TERU, University of Glasgow

Mack J, Lansley S (1985) *Poor Britain* Allen & Unwin, London

Martin D (1991) *Geographic Information Systems and their socioeconomic applications* Routledge, London

Molho I (1991) "Patterns and trends in local pay in Great Britain, 1975-76 to 1987-88" *Urban Studies* 28, 535-552

Moore B, Townroe P (1990) *Urban Labour Markets* HMSO, London

Morin J (undated) "La participation electorale" in A Fouquet, M Marpsat (eds.) *Donnees sociales 1987* INSEE, Paris

Morris R, Carstairs V (1991) "Which deprivation? A comparison of selected deprivation indexes" *Journal of Public Health Medicine* 13, 318-326

Newcastle upon Tyne City Council (undated) *Newcastle upon Tyne 'City Profiles' – results from the 1986 Household Suvey* Policy Services Dept., City of Newcastle upon Tyne

OPCS (1990) *Key population and vital statistics 1989: local and health authority areas* vs16/PP1-12, HMSO, London

OPCS & GRO(Scotland) (1992a) *1991 Census definitions* HMSO, London

OPCS & GRO(Scotland) (1992b) "Revised timetable for LBS and SAS output" *Census Newsletter* 19, 1-2

OPCS, GRO (Scotland) & CO (N.Ireland) (1992) *Statistical information on population and housing, 1996-2016: an invitation to shape future policy* UK Census Offices, London

Pilling D (1990) *Escape from disadvantage* Falmer, London

Richardson G, Grieve S, (1988) "Wages of sin" in M Goldring (ed.) *The world in 1989* Economist Publications, London

Rogerson R, Findlay AM, Morris AS, Coombes MG (1989) "Indicators of quality of life: some methodological issues" *Environment & Planning A* 21, 1655-1666

Savage D (1992) "Income question for use on the General Household Survey" *General Household Survey Newsletter* 10, 5-11

Sheldon E B, Moore W E (1968) *Indicators of social change: concepts and measures* Russel Sage, New York

Shucksmith M (1990) "The definition of rural areas and rural deprivation" *Research Report 2* Scottish Homes, Edinburgh

Smith DM (1979) "The identification of problems in cities: applications of social indicators" in DT Herbert & DM Smith (eds) *Social problems and the city* Oxford University Press, New York

Stewart J, Spencer K, Webster B (1974) *Local government and urban deprivation* Home Office Urban Deprivation Unit, London

Townsend P (1987) "Deprivation" *Journal of Social Policy* 16, 125-146

Townsend P, Davidson N, Whitehead M, Black D (eds.) (1988) *Inequalities in health: the Black report and the health divide* Penguin, Harmondsworth

Townsend P, Gordon D (1991) "What is enough? New evidence on poverty allowing the definition of a minimum benefit" in M Adler, C Bell, A Sinfield (eds.) *Sociology of social security* Edinburgh University Press, Edinburgh

Townsend P, Phillimore P, Beattie A (1988) *Health and deprivation: inequality and the north* Croom Helm: London

Training Agency (1990) *Inner cities: responding to the challenge* Developing good practice 3/1, Training Agency, Sheffield

United Nations Organisation (1954) *International definition and measurement of standards of living* Report by Committee of Experts to UN, New York

Walsh S, Barton S, Jarvis S, Clark W (1990) *Children injured in road traffic accidents: a collaborative project with Northumbria Police* NE.RRL Research Report 90/3, CURDS, University of Newcastle upon Tyne

Whitehead (1987) *The health divide: inequalities in health in the 1980s* Health Education Council, London

Wong C, Coombes M G, Raybould S, (1992) "The environmental quality of residential neighbourhoods: a town planning application of GIS" *Town Planning Review* 62, 369-373

Wormald P (1991) "The 1991 Census - a cause for concern?" *Population Trends* 66, 19-21

Wrekin Council Policy Unit (1991) *Wrekin Trends '91* Wrekin District Council, Telford

Appendix: technical details of the experimental analyses

Section 6 of the report explains that, in order to evaluate the effects of various methodologies empirically, a total of five methods were applied to a specially created dataset. As detailed in **Table A**, ten variables were created to provide the basis for the experimental analysis. The aim was to cover each of the Issues identified in the main report: in practice, no **Health** indicator could be developed so separate variables were collected for **Income** and **Basic Needs**. The variables were mostly drawn from the 1981 Census of Population, and designed to maximise the commonality of the data to both 1981 practice and the 1991 recommendations. In common with both, all the indicators were ratios, and none measured change through time. Practical limitations were, however, very important given the timescale for the project.

Data for Staffordshire was excluded due to technical problems with the source material, as were 'special EDs' (which have very low populations). This left a total of 99,660 cases. For parts of the analysis a 10% random sample was analysed.

The initial methodology applied was to re-create, as closely as possible, the 'Z scores' practice of 1981. Data were normalised if necessary, converted to Z scores and then the results for all indicators summed for each ED. For comparison at the level of the local authority District, the mean of the scores across EDs in the District was taken (not weighted by the population of the EDs). The analysis did not include the calculation of the share of each District's EDs in the "worst" 10% across the country (the final step of the Department's earlier targetting analysis based on 1981 data). Given the nature of the **Recreation** variable (for which deprivation is indicated by a low value) it was subtracted from the final score, rather than being added to it. This method then represents the Null weighting methodology.

The second methodology to be applied to the data was factor analysis. Principal Area Factoring was used to create one factor to 'explain' as much of the data variance as possible. In fact 24.6% of the variance was accounted for by this Single Factor, which weighted most heavily on the variables for **Employment, Communications, Housing** and **Education**.

Table A

Issue	Data source	DoE (1983) term	Definition ("c" = 1981 Census cell)
Social Conditions	1981 Census	'ethnic origin'	100 (c2920/c2875)
Physical Environment	1981 Census	'lacking basic amenities'	100 (c951-c952)/c951
Housing	1981 Census	'overcrowding'	100 (c1079 + c1080)/c951
Education	1981 Census	NEW VARIABLE	100 (c5154 + c5157 + c5158 + c5164)/ c5165
Employment	1981 Census	'unemployment'	100(c859/c719)
Work Conditions	Employment Census 1981	NEW VARIABLE	Interpolated from District level (% jobs in Activity Headings shown to be high risk by Annual Reports of Health and Safety Executive)
Income	1981 Census	NEW VARIABLE	100 (n + (0.5* (c639 + c640 + c641) - (c615 + c616 + c617))/n) where n = c1532 + c1533 + c1539 + c1540 + 1.7 (c1546 + c1547 + c1553 + c1554) + 2.7 (c1560 + c1561 + c1567 + c1568) + 0.5 (c2009 + c2010)
Basic Needs	1981 Census	NEW VARIABLE	100 (c2198 - (c2206 + c2221) + c2213)/ (c462 + c473)
Communication	1981 Census	NEW VARIABLE	100 (c2561/c2507)
Recreation	1981 Census	NEW VARIABLE	39.4 ((c640 + c641)/(c615 + c637 + c638 + c640 + c641 + c1878 - (c1881 + c1883 + c1885 + c1887 + c1888))) + 33.2 ((c637 + c638)/(c615 + c637 + c638 + c640 + c641 + c1878 - (c1881+ c1883 + c1885 + c1887 + c1888))) + 44.2 (c615/(c615 + c637 + c638 + c640 + c641 + c1878 - (c1881 + c1883 + c1885 + c1887 + c1888))) + 53.3 ((c1878 - (c1881 + c1883+ c1885+ c1887 + c1888))/(c615 + c637 + c638 + c640 + c641 + c1878 - (c1881 + c1883 + c1885 + c1887 + c1888)))

The third methodology was similar, except that an unconstrained number of factors were created using a Principal Components Analysis. These factors sought to account for all the data variance, but three accounted for the initial 54.2%. Here, the First Factor accounted for 29.6% of the variance. This factor was most heavily loaded on the same variables as the Single Factor.

Various methods of rotation were tried on the data. These are standard techniques to ensure that each Factor is independent of the others. While they made the individual factors somewhat easier to explain as syndromes, when aggregated to the District level the ranking were almost unchanged. Consequently, the basic unrotated factor solution was adopted as the more straightforward alternative.

All the first three methods were carried out within **SPSSx** Super-script. The final two methods required the use of customised FORTRAN programs, although the first of these used the output from the Principal Components Analysis. The first three factors were used as input because each of the 11 variables loaded (at a value of over 0.5) onto one, and only one.

The next step was to impose a simple form of Multi-Criteria analysis which can be visualised as a set of sieves, with one sieve for each factor. An ED that successfully passes through all three sieves is regarded as deprived. There are two levels of sophistication in addition to this. The first is the stringency of the sieves, and the second is the possibility that an ED can pass 'around' one of the sieves, but still be regarded as though it had passed through successfully.

Both of these features are controlled by the user, after inspection of the proportion of the population identified by different options. The density of the sieve is controlled by use of standard deviations: here EDs pass through the sieve if they have a Z score for that factor which is at least + 0.5 (ie greater than the mean for the factor by at least half of the standard deviation of all EDs' values). The second variable feature is the number of sieves an ED *must* pass. Adopting the setting of mean plus 0.5 standard deviations, and requiring that the ED passes at least two sieves, results in 17.5% of the economically active population being classified as deprived. This level of inclusion was felt to be appropriate because it would ensure that most sizeable areas had some of their population so classified: all plausible candidate areas would be recognised, and larger areas such as Districts could be assessed in terms of the share of their population included.

Cluster Analysis was adopted as the fifth and final method to be applied to the data. Specially modified programs were used in order to successfully analyse nearly 100000 cases. A range of solutions were run, giving numbers of clusters in the solution from three to fifty. Two clear clusters remained stable over the range of solutions up to around 10 clusters: the first was based upon the variables associated with **Physical Environment, Housing, Employment** and (particularly) **Social Conditions**, while the second was related to the variables for **Income, Communications, Education** and **Employment**.

The seven cluster solution was adopted as being the most appropriate here. The two clusters which were describable as 'deprived' between them contained 19.9% of the national total of the economically active population.